

**How Brazilian and Portuguese newsrooms use information  
visualization in journalism: a qualitative comparative analysis**

**Master's Dissertation  
in New Media and Web Practices**

**May, 2018**

Dissertation submitted in accordance with the requirements for the degree of Master of Arts in New Media and Web Practices, under scientific supervision of Prof. Dr. Ana Raquel de Ponte Figueiras.

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Novos Media e Praticas Web, realizada sob a orientação científica da Prof. Doutora Ana Raquel de Ponte Figueiras

*To all those who, at any point of my life,  
encouraged my international academic path.*

## ACKNOWLEDGEMENTS

The present dissertation is the outcome of a process that began a long time before the official start of this master's course. Several people have influenced such path, direct or indirectly. Some of them are very aware of that, others may not. Nominating them – with the evident risk of leaving someone out – is the way I found to thank for each contribution.

First, I offer my sincerest gratitude to my supervisor, Prof. Dr. Ana Figueiras, who guided me with patience and knowledge, while allowing me room to work in my own way. Her detailed revision and permanent support were fundamental for this work. I also would like to thank Prof. Paulo Nuno Vicente, my joint supervisor and iNOVA Media Lab coordinator. Appreciation to all members of the mentioned laboratory, where I spent a good part of my writing time. Here, I extend my greetings to the other professors of the course and to the professionals I have interviewed, including those I talked informally via the WhatsApp group *#ddjbr*. They all were very kind and considerate.

Impossible not to mention my family in Brazil. A special thanks to my father, mother and sister, who have always supported me in good and bad times, independently of the ocean who keeps us physically apart. My deep love to my nephew Dudu, for being a breath of fresh air when I feel stifled. I am also very grateful to my grandmother, Nelsy, who upheld me so many times in life and unfortunately passed away in the final instants of this academic path.

A friend deserves all my thankfulness and affection: Carol Delmazo, former classmate and flatmate and a forever sister of heart, always present in joy and pains. Carol introduced me to another special person, the cooperative and empathetic Alda Rocha, to whom I am also very grateful. Thanks to Richard Heffernan, ever looking after me. All my love to a special group of friends in Brasília, namely Morena, Priscila, Ana Carla, Morillo, Ricardo, Balbino e Caio, a real family that has taken a big part of my heart.

Finally, a big thank to Glademir Bressiani, Anderson Arcoverde, and all the CNT TV team for the partnership and trust placed in my work during all my time in Lisbon.



# **HOW BRAZILIAN AND PORTUGUESE NEWSROOMS USE INFORMATION VISUALIZATION IN JOURNALISM: A QUALITATIVE COMPARATIVE ANALYSIS**

**FERNANDA BARROS ABRAS**

## **ABSTRACT**

Data journalism is understood as a flow that encompasses at least four stages of data processing – gathering, filtering, analyzing and communicating – with the latter being developed preferably in digital media in order to include interactive visualizations. This practice has been led mainly by American and British publications and only recently incorporated by newsrooms in Brazil and Portugal. This study sought to investigate the field in both countries through a literature review, interviews with stakeholders and publications' assessment. It was verified that the Brazilian media outlets are ahead of the Portuguese, mainly because the companies in Brazil have realized the value of structuring multidisciplinary teams – comprising programmers and/or designers beyond journalists – fully dedicated to the data journalism practice. In Portugal, media outlets remain either resorting to only one data journalist – who is often absorbed by the daily hard-news – or not institutionally recognizing the data journalism practice.

**KEYWORDS:** data journalism, information visualization, narrative visualization, interactive visualization.

## **RESUMO**

O jornalismo de dados é entendido como um fluxo que inclui pelo menos quatro etapas de processamento de dados – coleta, filtragem, análise e comunicação – com o último estágio sendo desenvolvido em mídias digitais para que visualizações interativas possam ser incluídas. Essa prática tem sido liderada principalmente por publicações norte-americanas e britânicas e só recentemente vem sendo incorporada por redações no Brasil e em Portugal. Este estudo procurou investigar o campo de jornalismo de dados nos dois países por meio de uma revisão de literatura, entrevistas com as partes interessadas e avaliação de reportagens. Foi verificado que os veículos brasileiros estão à frente dos portugueses, principalmente porque as empresas de comunicação no Brasil perceberam o valor de estruturar equipes dedicadas exclusivamente ao jornalismo de dados que sejam multidisciplinares, incluindo programadores e/ou designers além de jornalistas. Em Portugal, as empresas jornalísticas ou têm apenas um jornalista de dados – que frequentemente é absorvido pela produção diária de notícias – ou não reconhecem institucionalmente a prática do jornalismo de dados nas suas redações.

**PALAVRAS-CHAVE:** jornalismo de dados, visualização de informação, visualização narrativa, visualização interativa.

## CONTENTS

<b>INTRODUCTION .....</b>	<b>1</b>
<b>1 – DATA JOURNALISM: A MULTIDISCIPLINARY PROCESS OF MAKING SENSE OF DATA ...</b>	<b>3</b>
1.1 – Big Data.....	3
1.2 – Concept, origins, and evolution of data journalism .....	5
1.3 – Information visualization: the data journalism workhorse .....	13
1.3.1 – Evolution of information visualization .....	14
1.3.2 – Information visualization literacy.....	20
1.3.3 – Narrative visualization: when storytelling meets Infovis .....	22
1.3.4 – Interactivity in information visualization .....	23
1.3.5 – Infovis tools: Tableau, D3.js, and Datawrapper .....	27
1.4 – Data stories production process: roles and skills .....	28
<b>2 – DATA JOURNALISM IN BRAZIL AND PORTUGAL .....</b>	<b>33</b>
2.1 – Methodology .....	33
2.1.1 – Choosing interviewees .....	35
2.1.2 – Choosing data stories .....	37
2.1.3 – Proposed taxonomy .....	39
2.2 – Brazilian historical context .....	41
2.3 – Portuguese historical context.....	45
<b>3 – FINDINGS.....</b>	<b>46</b>
3.1 – Interviews .....	46
3.2 – Volt Data Lab, Hemiciclo, and Frames.....	59
3.3 – Data stories qualitative evaluation .....	64
3.4 – Quantitative assessment of the DDJBR list .....	104
3.5 – Qualitative comparative analysis .....	107
<b>CONCLUSION.....</b>	<b>109</b>
<b>REFERENCES .....</b>	<b>112</b>
<b>APPENDIX 1: EXTRA FIGURES .....</b>	<b>i</b>
<b>APPENDIX 2: INTERVIEWS TRANSCRIPTIONS (IN PORTUGUESE) .....</b>	<b>xxii</b>

## LIST OF FIGURES, TABLES AND DIAGRAMS

Figure 1 - The Guardian revealed the rise in IED (improvised explosive device) attacks between 2004 and 2009, with an interactive map that allowed the audience to drag the date along a bar to see where and who they hit over these five years. ....	9
Figure 2 - The New York Times presented an interactive map to show the deaths on the 20th December 2006 in Baghdad. ....	9
Figure 3 - Der Spiegel produced an interactive graphic which selected events plotted on a map, and a historical timeline chart showing the deaths and woundings in the context of key historical milestones. ....	10
Figure 4 - Data journalism process proposed by Lorenz (2010).....	11
Figure 5 - William Playfair chart, The Commercial and Political Atlas, 1786 .....	15
Figure 6 - Manchester Guardian, 5th May 1821 (full version in Appendix 1) .....	16
Figure 7 - Line chart showing the magnitude of cholera deaths in New York ("New-York Daily Tribune., September 29, 1849, Image 1," n.d.).....	16
Figure 8 - John Snow's cholera map of Soho, 1854.....	17
Figure 9 - Florence Nightingale's Rose Diagram, 1858.....	18
Figure 10 - Charles Minard's chart, 1861: losses of Napoleon's army in Russia, 1812-1813.....	19
Figure 11 - This is an American Workday, By Occupation (Yau, 2017) .....	21
Figure 12 - Visual Data Storytelling Process (Lee et al., 2015) .....	30
Figure 13 - Proposed hierarchical taxonomy .....	40
Figure 14 - The first Estadão Dados' data-driven project: Basômetro .....	42
Figure 15 - The slideshow animation is built with the 16 maps below .....	66
Figure 16 – The sequence shows the evolution of MC Kevinho is one of the three sets. The two other sets of map and line chart were not shown as they are very similar.....	66
Figure 17 – The section "busca" is the only interactive part. The example below shows the popularity of Michael Jackson in Brazil, a picture of him and the YouTube videoclip of the hit "Thriller" .....	67
Figure 18 - The story has only a static map and two bar charts .....	69
Figure 19 - First view of Estadão's narrative visualization .....	70
Figure 20 - Animation shows the stolen cars in São Paulo between 2008 and 2017. Below, 2008 and 2013 maps are represented.....	71
Figure 21 - The first screen shows a detail of the map with an explicative text. The second presents a photo and a quotation of a man who had the car stolen, and, in the background, the map with the exact point where the theft happened. ....	72
Figure 22 - The story climax: a browsable map.....	73
Figure 23 - G1's story presents four drop-down menus and one tab that leads to an annual search .....	74
Figure 24 - The annual data of violent deaths in Brazil.....	75
Figure 25 - Nexo's story first view, with a short introductory text and search boxes .....	76
Figure 26 - A bar chart shows the city's expenses by field of application .....	77

Figure 27 - Three examples of the compound graphics that show data of each field of budget application.....	77
Figure 28 - Comparison between cities presents reconfiguration options.....	78
Figure 29 - Interactive graphic with drop-down menu .....	79
Figure 30 – The interactive chart shows the relation between items when hovering the cursor on them. ....	80
Figure 31 - An interactive map shows the distribution of criminality by criminal factions and militia groups in the Rio de Janeiro State. ....	80
Figure 32 - Graphics made with Tableau: the interactive task is detail-on-demand .....	81
Figure 33 - The chart can be reconfigured when clicking on "keep only" or "delete" .....	82
Figure 34 - Two interactive slides of the sequence that tells the story of Pedro Dias.....	83
Figure 35 - The scrollytelling creates an effect of interactive sequence, with the text running across the screen over the graphic background .....	85
Figure 36 - The three interactive tasks: details-on-demand, relate and zoom.....	86
Figure 37 - Each tab shows a reconfigured bar chart.....	88
Figure 38 - O Globo's story resorts to a compound static graphic.....	89
Figure 39 - Estadão's story presents a compound static infovis .....	90
Figure 40 - A Sankey diagram representing judicial magistrates' wages! (this illustration was shortened).....	91
Figure 41 - Other kinds of charts used by Nexo .....	91
Figure 42 - Volt data Lab's static compound visualization about the judicial salaries in São Paulo state.....	92
Figure 43 - Example of the interactive task details-on-demand.....	93
Figure 44 - Details-on-demand and zoom are the interactive tasks in this story .....	94
Figure 45 - This Público's story resorts only to static graphics and map .....	95
Figure 46 - Público's interactive story produced under the REACTION project.....	97
Figure 47 - The interactive tasks are zoom and details-on-demand.....	98
Figure 48 - Example of a 'frame', its sharing options and box for suggestions.....	99
Figure 49 - Hemiciclo's landing page.....	100
Figure 50 - Interactive stacked bar chart, with drop-down menus .....	100
Figure 51 - Other interactive charts on the Hemiciclo's landing page .....	101
Figure 52 - Visualization of a voting: in the left, the colored dots show who voted for and in the right, the colored dots show the abstentions.....	102
Figure 53 - Data stories distribution by media outlet .....	104
Figure 54 - Data stories distribution by media outlet and month .....	105
Figure 55 - Data stories distribution by the presence or absence of visualizations .....	105
Figure 56 - Stories with visualizations .....	106
Figure 57 - Stories with dynamic visualizations .....	107

Figure 58 - Stories with static visualizations .....	107
Figure 59 - Manchester Guardian, 5th May 1821 .....	i
Figure 60 - Explanation of how to read the line chart that shows the magnitude of cholera deaths in New York ("New-York Daily Tribune., September 29, 1849, Image 1," n.d.) .....	ii
Figure 61 - Overview of "Música Muito Popular Brasileira" (Folha): due to its extension, the story was divided in 5 columns .....	iii
Figure 62 - Overview of "Barcelona, capital catalã, teve maior proporção de votos contrários à separação" - O Globo .....	iv
Figure 63 - Overview of "As mortes violentas mês a mês no país" (G1).....	v
Figure 64 - Overview of "Como sua cidade gasta dinheiro? Descubra e compare com outras" (Nexo): each column represents possible configurations, according to the filter applied. The first shows information about a city. The second is a comparison between two cities. ....	vi
Figure 65 – Overview of "Tá tudo dominado" - (Volt Data Lab for The Intercept Brasil): due to its extension, the story was divided in 3 columns .....	vii
Figure 66 - Overview of "A nação enquanto não há remodelação: 12 gráficos para explicar o bom e o mau do país" (Expresso): due to its extension, the story was divided in 3 columns.....	viii
Figure 67 - Overview of "Em 20 anos, o que mudou em Portugal?"(Público): due to its extension, the story was divided in 3 columns .....	ix
Figure 68 - Overview of "Mapa interativo. Sabe o significado literal do nome dos países? (Observador) .....	x
Figure 69 - Overview of "Benefícios a juízes nos Estados têm disparidades e distorções (Folha): due to its extension, the story was divided in 2 columns .....	xi
Figure 70 - Overview of "Com extras, 71% dos juízes do país recebem acima do teto de R\$ 33 mil" (O Globo) .....	xii
Figure 71 - Overview of "TJs gastam R\$ 890 mi com 'penduricalhos' para juízes" (Estadão) .....	xiii
Figure 72 - Overview of "Quanto ganham desembargadores e juízes no Brasil" (Nexo): due to its extension, the story was divided in 2 columns .....	xiv
Figure 73 - Overview of "Togas de ouro" (Volt Data Lab for Piauí Magazine) .....	xv
Figure 74 - Overview of "A estrada da morte: onde o fogo matou – e onde ainda lavra (mapa interativo)" (Expresso).....	xvi
Figure 75 - Overview of "As freguesias nas listas de prioridade de limpeza de terrenos" (Jornal de Notícias).....	xvii
Figure 76 – Overview of "Incêndio de Pedrógão Grande é o maior de sempre em Portugal" (Público) .....	xviii
Figure 77 - Overview of "Doze anos de incêndios" (Público).....	xix
Figure 78 - Overview of "Veja o mapa interativo dos incêndios no centro do país" (Observador) .....	xx
Figure 79 - Overview of "Área ardida em Portugal" - Frames (for Observador).....	xxi
Table 1 - Matrix of Interactivity (Jensen, 2008) .....	24
Table 2 - Interactivity taxonomies by author .....	25

Table 3 - Circulation of leading Brazilian newspapers with data journalism production .....	34
Table 4 - Circulation of leading Portuguese newspapers with data journalism production.....	34
Table 5 - Data journalism professionals surveyed .....	36
Table 7 - Snapshot of data journalism structure in the surveyed outlets.....	59
Table 8 - Summary of "Música muito popular brasileira", Folha .....	68
Table 9 - Summary of "Barcelona, capital catalã, teve maior proporção de votos contrários à separação", by O Globo.....	69
Table 10 - Summary of "Em queda, roubos de carros migram para a periferia", by Estadão ....	73
Table 11 - Summary of "As mortes violentas mês a mês no país", by G1.....	75
Table 12 - Summary of "Como sua cidade gasta dinheiro? Descubra e compare com outras", by Nexo .....	78
Table 13 - Summary of "Tá tudo dominado", by Volt Data Lab (for The Intercept Brasil) .....	81
Table 14 - Summary of "A nação enquanto não há remodelação: 12 gráficos para explicar o bom e o mau do país", by Expresso .....	82
Table 15 - Summary of "O mapa interativo da fuga do homem mais procurado em Portugal", by Jornal de Notícias .....	84
Table 16 - Summary of "Em 28 anos, o que mudou em Portugal?", by Público .....	85
Table 17 - Summary of "Mapa interativo. Sabe o significado literal do nome dos países?", by Observador.....	87
Table 18 - Summary of "Benefícios a juízes nos Estados têm disparidades e distorções", by Folha.....	88
Table 19 - Summary of "Com extras, 71% dos juízes do país recebem acima do teto de R\$ 33 mil", by O Globo .....	89
Table 20 - Summary of "TJs gastam R\$ 890 mi com 'penduricalhos' para juízes", by Estadão ..	90
Table 21 - Summary of "Quanto ganham desembargadores e juízes no Brasil", by Nexo .....	92
Table 22 - Summary of "Togas de ouro", by Volt Data Lab (for Piauí Magazine) .....	93
Table 23 - Summary of "A estrada da morte: onde o fogo matou – e onde ainda lavra (mapa interativo)", by Expresso .....	94
Table 24 - Summary of "As freguesias nas listas de prioridade de limpeza de terrenos", by Jornal de Notícias .....	95
Table 25 - Summary of "Incêndio de Pedrógão Grande é o maior de sempre em Portugal, by Público.....	96
Table 26 - Summary of "Doze anos de incêndios", by Público.....	97
Table 27 - Summary of "Veja o mapa interativo dos incêndios no centro do país", by Observador.....	98
Table 28 - Summary of "Área ardida em Portugal", by Frames (for Observador) .....	99
Table 29 - Summary of Hemiciclo.....	102
Table 30 - Summary of the qualitative analysis .....	103

## INTRODUCTION

---

*Data-driven journalism is the future. Journalists need to be data-savvy. It used to be that you would get stories by chatting to people in bars, and it still might be that you'll do it that way sometimes. But now it's also going to be about poring over data and equipping yourself with the tools to analyze it and picking out what's interesting.*

*Tim Berners-Lee (as cited in The Data Journalism Handbook)*

---

Some might argue that the founder of the World Wide Web was too geeky when he made the above declaration. It was in 2010, during the launch of the first UK government spending datasets. Nevertheless, if Tim Berners-Lee said that today many would say that he was completely right. Indeed, almost a decade ago very few journalists were concerned with processing, analysis, and communication of data. On the other hand, nowadays newsrooms have increasingly been training and hiring human resources to incorporate data-savvy professionals to their staffs.

Even so, data journalism remains incipient worldwide. The United States', United Kingdom', and Germany's media outlets have been taking the lead in the last decade, publishing flagship data journalism stories with sophisticated interactive visualizations. Only recently this practice has gained momentum in other countries, such as Brazil and Portugal. A brief glance at both countries' publications reveals that some vehicles have been incorporating sparse initiatives with embryonic attempts of using information visualizations. Likewise, the related academic production in the two countries is scarce.

Therefore, there is a need for further recent and in-depth studies to characterize the data journalism state-of-the-art in Portugal and Brazil, as well as to propose guidelines to facilitate the use of datasets as the main raw material of visual stories. Hence, this study sought to investigate the field through a literature review, interviews with stakeholders and publications' assessment. The goal was to answer two principal investigation questions, a broader and a more specific.

*Q1: What is the current status of data journalism in Brazil and Portugal?*

*Q2: How information visualization is being used in the context of journalism in both countries?*

The final aim was promoting a cross-national comparison by means of a Qualitative Comparative Analysis (QCA), as a layer on top of a multiple-case study. The Brazilian and Portuguese media outlets were chosen considering their relevance in a national scale and the practice of data journalism, even if incipient. In Brazil, were analyzed three online versions of traditional newspapers ([O Estado de S. Paulo](#), [Folha de S. Paulo](#), and [O Globo](#)), two online-only news publications ([G1](#) and [Nexo](#)) and one independent initiative ([Volt Data Lab](#)). In Portugal, were investigated three online versions of traditional newspapers ([Expresso](#), [Jornal de Notícias](#), and [Público](#)), one online-only news publication ([Observador](#)) and one independent initiative ([Hemicycle](#)).

Three methodological approaches were applied to gather the information: semi-structured interviews with data journalism professionals from all surveyed media outlets, a qualitative evaluation of data stories produced by these vehicles, and a quantitative assessment of a recent list created by Volt Data Lab with the best Brazilian data journalism stories, chosen by Brazilian journalists from various media companies.

This dissertation is divided into three chapters. The first is a bibliography review that comprises a contextualization and conceptualization of data journalism, with emphasis on information visualization as the more appropriated way for communicating stories built with data. The second chapter brings the methodology and the historical context of data journalism in Brazil and Portugal. Finally, the third chapter details the investigation's results.



## 1 – DATA JOURNALISM: A MULTIDISCIPLINARY PROCESS OF MAKING SENSE OF DATA

---

*When information was scarce, most of our efforts were devoted to hunting and gathering. Now that information is abundant, processing is more important.*

*Philip Meyer (as cited in The Data Journalism Handbook)*

---

The “increasingly ubiquitous digitization of information” (Coddington, 2015, p.331), combined with open source software and freedom of information culture, spawned the need of bringing sense and structure out of the endless flow of data. For journalism, it represents an opportunity to get valuable information and present relevant stories to the audiences. Adding practices, skills, and even professionals from other fields – statistics, design, and programming, for instance – to the traditional journalism workflow was necessary, in a process that includes gathering, selection, analysis, interpretation, and presentation of datasets. Such convergence represents an expansion of the press role as an advocate of the public interest.

### 1.1 – Big Data

These days, it is fashionable to talk about big data. The buzzword is everywhere and has been defined in different ways. Hence, it is often misunderstood. Dan Ariely<sup>1</sup>, author and professor of Psychology and Behavioral Economics, on the 6th January 2013, posted on Facebook: “big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...”<sup>2</sup>. Over five years later, the joke is still valid.

The designation has been in vogue ever since, and most dictionaries already come up with definitions for it. According to the Cambridge Dictionary, big data is “very large sets of data that are produced by people using the internet, and that can only be

---

<sup>1</sup> <http://danariely.com>

<sup>2</sup> <https://www.facebook.com/dan.ariely/posts/904383595868>

stored, understood, and used with the help of special tools and methods”<sup>3</sup>. The Oxford English Dictionaries state that it is a term of the computing field, which means “extremely large data sets that may be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions”<sup>4</sup>.

The term big data was firstly used in the academic sphere, before hitting the broader public and commercial realm. It emerged two decades ago, in a paper by the NASA researchers Michael Cox and David Ellsworth (1997), to describe the problem of computers generating large amounts of data that cannot be processed and visualized.

Visualization provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of *big data* (Cox and Ellsworth, 1997, p.235, emphasis added).

From then on, the idea of big data has spread to all sectors of modern life. The developments of communications technology have increasingly interconnected people and things. “Generally referred to as machine-to-machine (M2M), interconnectivity is responsible for double-digit year over year (YoY) data growth rates. Finally, because small integrated circuits are now so inexpensive, we’re able to add intelligence to almost everything” (Zikopoulos et al., 2012, p.4).

Such massive amount of data has been used by the most diverse segments – such as business, technology, media, science and medicine – to make analysis, correlations and predictions, with standard software capable of managing and manipulating data sets that once required supercomputers (Manovich, 2012; Zikopoulos et al., 2012).

In search of a definition, authors cite the *V’s of big data* (Lewis and Westlund, 2015; Mayer-Schönberger and Cukier, 2013; Stone, 2014; Zikopoulos et al., 2012):

- **Volume:** whilst “little” data storage is measured in up to gigabytes and fit on a personal computer, big data can only be stored on the cloud or other big storing systems, and measured in terabytes, petabytes, zettabytes and beyond.

---

<sup>3</sup> <https://dictionary.cambridge.org/dictionary/english-portuguese/big-data>

<sup>4</sup> [https://en.oxforddictionaries.com/definition/big\\_data](https://en.oxforddictionaries.com/definition/big_data)

- **Variety:** big data includes raw, semi-structured and unstructured data from web pages, web log files, search engines, social media, forums, email, documents, sensor data and so four.
- **Velocity:** it is related to the high growth rates of data, but also to the speed at which data is flowing.
- **Value:** large amounts of data reveal patterns and trends, from which is possible to extract valuable insights and revenue.

Therefore, day by day, all of us are contributing – like it or not – to the already vast informational ocean, so it is quite understandable why the subject is so vaunted. “Big data is poised to reshape the way we live, work, and think” (Mayer-Schönberger and Cukier, 2013). Indeed, it has impacted many industries and professions, including media and journalism.

Newsrooms have been training and hiring human resources, as well as using tools to simplify the data analysis process (Stone, 2014), leading to the emergence of the sub-field *data journalism*, which “can be understood as journalism’s response to the datafication of society” (Loosen et al., 2017, p.2). Some have described the big data era as the condition for data journalism’s practice. However, this is an erroneous assumption since the data journalists’ usual work rarely deals with petabytes or exabytes of data. Data journalism is much more related to quality than quantity: what matters is *good* data, which is not necessarily *big* data. What can be said is that the notion of big data has the potential to place data journalism in a higher threshold, as a catalyzer for a new kind of journalism that has its roots in older practices.

## 1.2 – Concept, origins, and evolution of data journalism

Conceptualizing data journalism can be quite an audacious job, since the field has experienced a significant growth in the last years and, except for a few cases, its practice is still incipient worldwide. Loosen et al. analyzed projects nominated for the Data Journalism Awards from 2013 to 2016 and noted that the number of countries represented by the nominees grew each year, what suggests that data journalism is progressively spreading around the world. However, the authors underline that projects

from the United States and from the United Kingdom constitute the majority, “likely influenced by the fact that data journalism has a longer history in anglophone countries (...)” (Loosen et al., 2017, p.14).

Another drawback is the use of many terms, coined in different contexts, as synonyms. After reviewing the related literature, Mark Coddington (2015, p.2) observed that the variety of terms “have produced a cacophony of overlapping and indistinct definitions that forms a shaky foundation for deeper research into these practices”. The author suggests that it is important to highlight their differences to build a consistent body of research. Hence, Coddington classified data-driven approaches to journalism in three categories, which appears to be a sensible taxonomy: computer-assisted reporting, computational journalism, and data journalism<sup>5</sup>.

*Computer-assisted reporting* (CAR) is the antecessor of both computational journalism and data journalism. It dates back to 1952 when the TV network CBS used a mainframe computer – a Remington Rand UNIVAC – and statistical models to predict the results of the presidential elections. In the late 1960s, the journalist Philip Meyer, one of the CAR’s pioneers, was granted a Nieman Fellowship at Harvard University to study quantitative methods used in social science. Meyer proposed applying these methods to journalism using computers and programming for statistical analysis, and renamed CAR as *precision journalism*, which became closely tied to investigative reporting (Bounegru, 2012; Coddington, 2015; Howard, 2014; Knight, 2015).

*Computational journalism* can be understood as a technologically oriented strand that comprises information gathering, organization, sense-making, presentation and dissemination (Coddington, 2015; Diakopoulos, 2011). An example would be the use of algorithms producing the news content itself, with low levels of human involvement. This category seems to fit better in the big data context, as it is quite impossible for human beings to analyze and process huge deluges of information without specific software.

At last, *data journalism*, the category approached in this dissertation, is still a concept under construction. The term is generally used by the news industry for

---

<sup>5</sup> Some authors also use the term data-driven journalism.

journalism practices based on data analysis, although professional definitions have tended to be broad and more focused on the *how* rather than the *why* to do it. The Guardian's journalist Paul Bradshaw (Bradshaw, 2010) stated that data journalism "represents the convergence of a number of fields which are significant in their own right - from investigative research and statistics to design and programming". Two years later, Bradshaw argues that *data* and *journalism* are troublesome terms, as journalists have always worked with a collection of numbers. He tried to clarify what makes data journalism different from the rest of journalism, arguing that it is a combination of the traditional nose for news and the ability to tell an appealing story using the big amounts of digital information available nowadays (Bradshaw, 2012). However, this is still not enough to distinguish data journalism. Liliana Bounegru (2012) goes beyond. The author claims that, while datasets were used to leverage the journalistic reportage since the end of the twentieth century, "data journalism pays as much — and sometimes more — attention to the data itself, rather than using data simply as a means to find or enhance stories". Thus, data journalism can be understood as a 21<sup>st</sup> century phenomenon, despite being an inheritor of previous journalistic practices, such as computer-assisted reporting and precision journalism.

The first to mention the data journalism practice – without naming it – was the software developer and journalist Adrian Holovaty (2006). In a blog entry, that later would be recognized as the data journalism manifesto<sup>6</sup>, he argued that "newspapers need to stop the story-centric worldview" and suggested the idea of telling stories in a structured format, using structured information.

So much of what local journalists collect day-to-day is structured information: the type of information that can be sliced-and-diced, in an automated fashion, by computers. Yet the information gets distilled into a big blob of text -- a newspaper story -- that has no chance of being repurposed (Holovaty, 2006).

---

<sup>6</sup> Some authors refer to it as database journalism manifesto, others as data-driven journalism manifesto.

However, the trailblazer in using the literal term *data journalism* seems to be Simon Rogers<sup>7</sup>, in December 2008, in a post on The Guardian Insider Blog. Roger came up with a large zip file with loads of spreadsheets about unemployment in the UK, available on the Office for National Statistics website, and wrote:

For us, this stuff can be a nightmare — to draw a map the graphic artist has to individually draw each constituency and colour it, which can take hours. Or used to. As of yesterday, our development team has come up with an application which takes the raw data and turns it into an editable map. Which meant that we could produce a fantastic interactive graphic based on these figures. It's *data journalism* — editorial and developers producing something technically interesting and that changes how we work and how we see data (Rogers, 2008, emphasis added).

Three months later, Rogers launched The Guardian Datablog. Using structured data published on the UK's Parliament website<sup>8</sup>, the British newspaper presented an infographic with all Members of Parliament (MP)'s spending, launching a public debate that has become known as MPs expenses scandal. The Guardian journalists relied on readers' help to scrutinize 450,000 documents. After that, the British government started releasing enormous amounts of data every year (Rogers, 2012).

It is clear that the open data, open source software and transparency culture of the mid 2000s has had an impact on data journalism rising (Coddington, 2015; Howard, 2014). Rogers (2012) highlights a turning point for the availability of public data: the then United States President Barack Obama opening of the US government's data faults<sup>9</sup> as his first legislative act, being followed by government data sites around the world.

However, what made data journalism notorious was the release of the Afghanistan and Iraq War Logs in 2010 by WikiLeaks. Julien Assange, the founder of the whistleblowing website, selected The New York Times, The Guardian, and Der Spiegel (German weekly magazine) to access and make sense of 483,000 records produced by the US Army – 92,000 of the Afghanistan war and 391,000 of the Iraq war. The three

---

<sup>7</sup> Today, Simon Rogers is data editor at Google.

<sup>8</sup> [www.parliament.uk/about/mps-and-lords/members/pay/mps/](http://www.parliament.uk/about/mps-and-lords/members/pay/mps/)

<sup>9</sup> [www.data.gov](http://www.data.gov) is a U.S. government website launched in late May 2009.

media outlets turned the complex data in interactive graphics and maps (figures 1, 2 and 3), telling detailed stories in a visual format (Howard, 2014; Kirk, 2010; Knight, 2015). If the same information had been published as text only, surely the result would have been a long article with much less richness of details and appeal.

Figure 1 - The Guardian revealed the rise in IED (improvised explosive device) attacks between 2004 and 2009, with an interactive map that allowed the audience to drag the date along a bar to see where and who they hit over these five years.

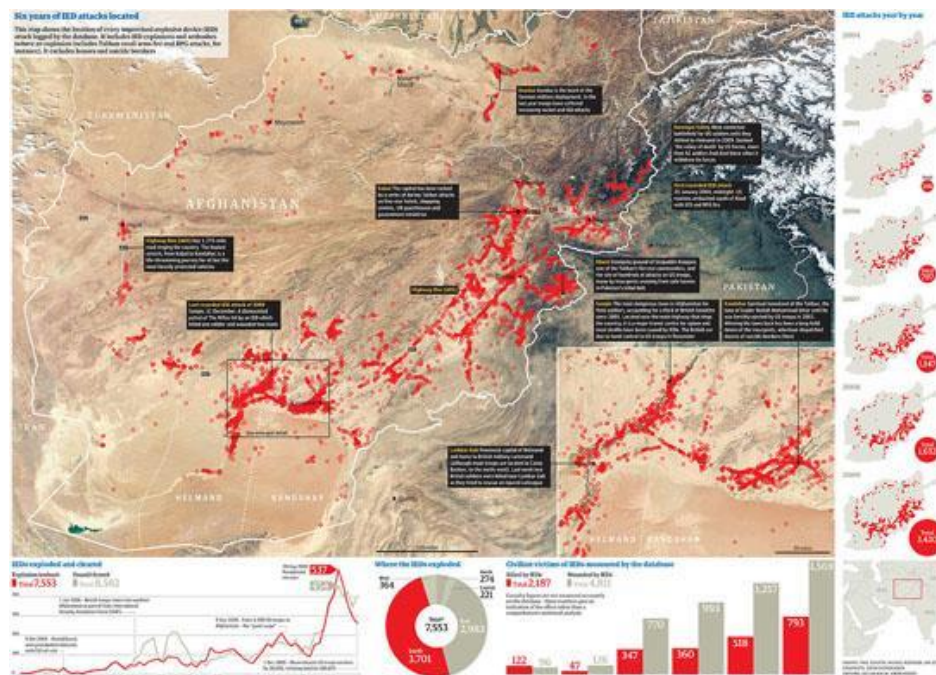
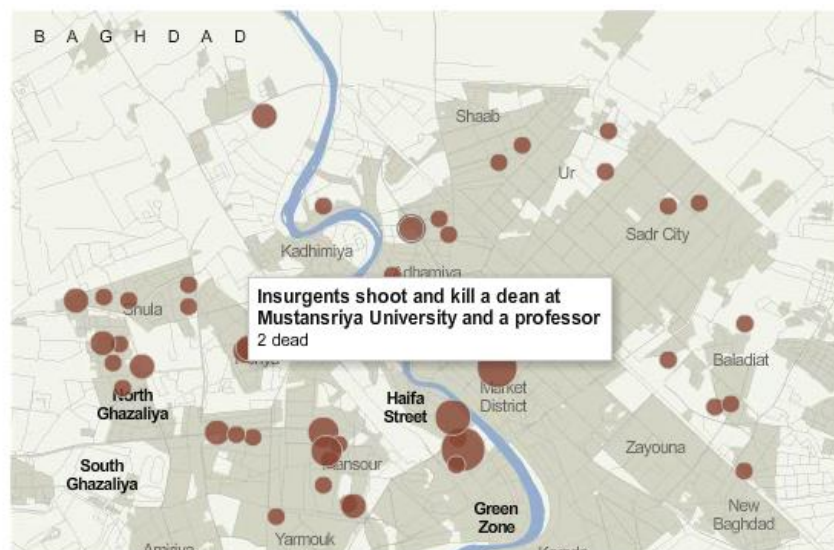


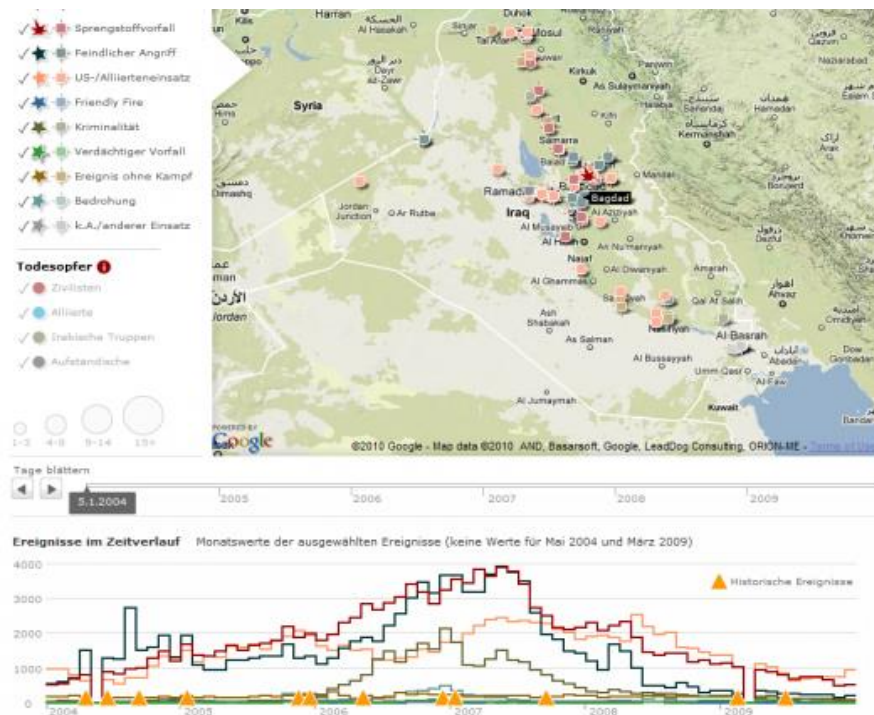
Figure 2 - The New York Times presented an interactive map to show the deaths on the 20th December 2006 in Baghdad.

## A Deadly Day In Baghdad

Violence peaked in December 2006, just two months before American troops arrived as part of what was later called “the surge.” At right are the details of one of the city’s deadliest days, Dec. 20. There were 114 separate episodes of violence that day, resulting in the deaths of about 160 Iraqi citizens and police officers.



**Figure 3** - Der Spiegel produced an interactive graphic which selected events plotted on a map, and a historical timeline chart showing the deaths and wounding in the context of key historical milestones.



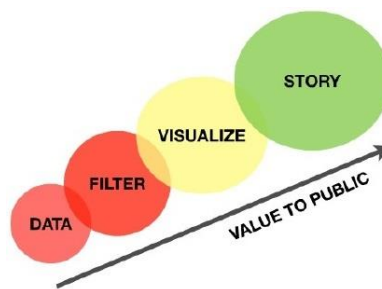
Rogers (2011) highlights that “Wikileaks has been the game-changer for data journalism”, pushing its adoption among newsrooms. Moreover, Wikileaks Afghanistan and Iraq stories, by The New York Times, The Guardian, and Der Spiegel, consolidated two fundamental elements of data journalism: information visualization and interactivity. Veglis and Bratsas included both factors in their definition of the practice:

(...) the process of extracting useful information from data, writing articles based on the information and embedding visualizations (interacting in some cases) in the articles that help readers to understand the significant of the story or allow them to pinpoint data that relate to them(Veglis and Bratsas, 2017,p.111).

Veglis and Bratsas also suggest that data journalism is a process, as well as Lorenz (2010), who called it “a process of refinement, where raw data is transformed into something meaningful”. According to Lorenz, the value to the public grows as datasets are filtered, visualized and turned into a clear story that people can easily understand and remember (figure 4).



Figure 4 - Data journalism process proposed by Lorenz (2010)



Data journalism workflow is described in slightly different ways by researchers (Bradshaw, 2012; Lorenz, 2010; Rogers, 2011; Veglis and Bratsas, 2017), but the essence tends to be the same, as presented below:

- The first stage – **gathering** – starts either with a question that needs data or a dataset that needs questioning. Data can be supplied by a variety of sources: directly by an organization, sometimes under open data legislation; through specific tools for advanced searching, scraping hidden databases and converting documents into something that can be analyzed; or simply through observation, surveys, online forms or crowdsourcing.
- **Filtering** data, also known as scrubbing or cleaning, means removing human error, flout unnecessary information, correcting corrupted records, tidying up and converting all the data into the same format.
- **Analyzing** datasets means contextualizing, investigating its origins and motivations, validating them, making calculations, and, if necessary, merging with further data.
- The last stage is **communicating** the story. Storytelling can take place through text, maps, tables, charts, infographics, animations and so four. As mentioned before, at this stage, *interactivity* and *visualization* have a fundamental role. The amount of text can vary considerably according to the subject, audience, and platform (mobile or desktop). In some cases, the visualization appears as a supplement of the text, but, as much as possible, the visualization should act as the center of the storytelling, with text playing a supplemental role.

Therefore, for the purpose of this dissertation, data journalism, **at its best**, would include all the stages cited above, with the latter being developed in digital media<sup>10</sup>, so that interactivity and visualization could be explored in all their potential to present stories based on datasets.

It does not mean journalists must adopt visualizations uncritically. Technological innovations in journalism might be seen not as a goal, but a way to enhance the practice of getting the message through. It seems that publishers have realized data visualization's value, as reveals the research led by Alberto Cairo (2017a). The author analyzed the winners of more than two decades of the annual Malofiej International Infographics Awards<sup>11</sup> and observed a changeover – due to the rise of digital data and technologies – from pictorial schemes and explanations to data-driven graphics based on abstract charts, diagrams, and maps. According to Cairo, this shift caused uneasiness to some, such as Javier Errea, the principal organizer of the Malofiej competition, when discussing the winners of the 2012 edition:

It's not just that data visualization is trendy among infographics designers. It's that its presence is overwhelming [...] Visualizations almost never inform, at least not clearly. Even worse: they make you feel guilty. If you don't understand them, it's because you're dumb. (Errea *apud* Cairo, 2017a, p.16)

Indeed, due to its easy-to-go-viral nature, visualizations are sometimes overused and, worse, misused. In addition, the trendy aspect, so common among technological developments, can undermine its job of communicating an idea better than the raw data itself could do. For this reason, ethical and professional standards remain as essential guidance. The principles of journalism – accuracy, truth, independence, fairness, humanity, accountability, etc. – are fundamental elements to guarantee that data

---

<sup>10</sup> Print media can also have data journalism teams, but the last stage of the work remains incomplete, since the visualization is static.

<sup>11</sup> The Malofiej Awards takes place in Pamplona, Spain, since 1993. It is considered the most influential and prestigious competition in the field of news graphics published in print and on-line across the globe. After every edition, the Malofiej organization publishes a book showcasing the winners.

visualization plays its role of conveying information in ways that are more effective and pleasing to the eyes of the audience. As Sarah Cohen (2012) highlights:

Not only can it be strikingly beautiful and attention-getting — valuable social currency for sharing and attracting readers — it also leverages a powerful cognitive advantage: fully half of the human brain is devoted to processing visual information. When you present a user with an information graphic, you are reaching them through the mind’s highest-bandwidth pathway. A *well-designed* data visualization can give viewers an immediate and profound impression, and cut through the clutter of a complex story to get right to the point (Cohen, 2012, emphasis added).

As the data journalism literature reveals, a few publications in the mainstream media have been successfully implementing the whole data journalism workflow, with well-designed interactive visualizations. The most prominent are The Guardian and The New York Times.

### 1.3 – Information visualization: the data journalism workhorse

Before being appropriated by different fields – such as human-computer interaction, computer science, graphics, visual design, engineering, statistical modeling, marketing, etc. – the definition of visualization was nothing more than what we usually find on dictionaries: the formation of a mental image of something<sup>12</sup>.

However, the old term has gained a new approach, and even dictionaries have been adopting the novelty. Now, visualization can also be “a chart or other image that is created as a visual representation of an object, situation, or set of information”<sup>13</sup>. The latest addition to the dictionary word definition is, actually, very close to many of those found in the academic literature for *information visualization* (or simply *Infovis*)<sup>14</sup>. Two decades ago, Gershon, Eick, and Card wrote:

---

<sup>12</sup> Source: The Oxford English Dictionaries

<sup>13</sup> Source: The Oxford English Dictionaries

<sup>14</sup> The term data visualization (or simply *datavis*) is also used interchangeably, as the definitions of information and data often overlap. As much as possible, we refer to the practice as information visualization, but the media industry prefers data visualization.

Visualization is the process of transforming data, information, and knowledge into visual form making use of humans' natural visual capabilities. With effective visual interfaces we can interact with large volumes of data rapidly and effectively to discover hidden characteristics, patterns, and trends (Gershon et al., 1998, p.9).

The most widely used definition of the term was given by Card, Mackinlay, and Shneiderman (1999, p.6, emphasis added): “the use of computer-supported, *interactive*, visual representations of abstract data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space”. As we can notice, the authors included the notion of interactivity within the concept, which makes sense, as the popularization of the internet and computer technology were the main factors for the development of the research in the field of information visualization.

However, in the data journalism research and industry, the term visualization appears, most of the time, with no explicit mention or incorporation of interactivity. Thus, its generally used interchangeably with infographic, graph, map, chart, diagram, etc., to describe any visual display of information and/or data. Because of that, in this dissertation, interactivity is approached as a layer on top of information visualization (or infographic, graph, map, chart, diagram), as can be seen in greater depth further on.

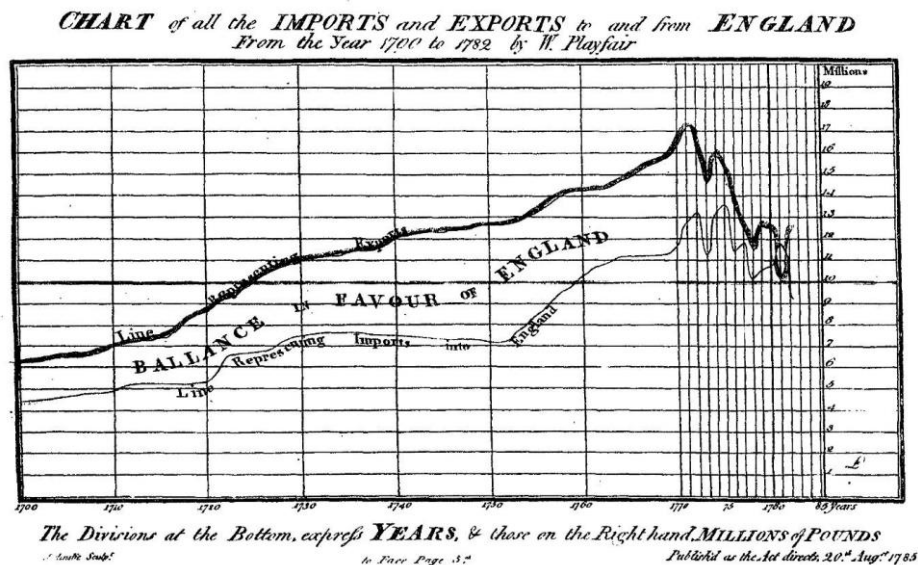
### ***1.3.1 – Evolution of information visualization***

The ability of the human brain to process visual information has been used since the early days of humankind. The oldest cave paintings and the Egyptian hieroglyphics could easily be called the first forms of information visualization (Krum, 2014). We can also find examples of some graphs, timelines, and maps in the 1500s and 1600s. However, with the rising of the probability theory and demographic statistics, the 18<sup>th</sup> century was the period that “witnessed, and participated in, the initial germination of the seeds of visualization that had been planted earlier” (Friendly, 2005, p.4).

The Scottish engineer William Playfair is widely considered the father of infographics, as he was the inventor of most of the graphical forms used today – line graphs, bar chart, pie charts and circle charts (Friendly, 2005). In 1786, Playfair published

*The Commercial and Political Atlas*, which displayed many time-series charts, representing the economy in England (Cairo, 2017b; Sachs, n.d.; Tufte, 1983). The first graph, for example, shows the totals of England's imports and exports throughout the eighteenth century (figure 5).

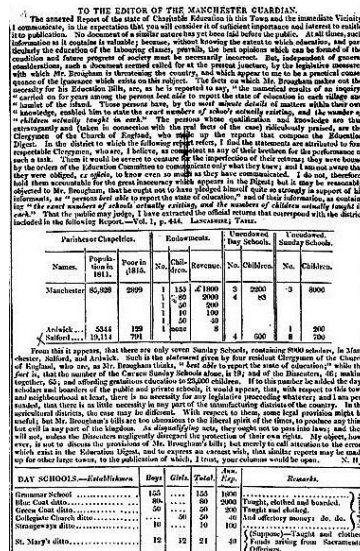
*Figure 5 - William Playfair chart, The Commercial and Political Atlas, 1786*



The next hundred years was time for infographics to take off. Especially the second half of the 19<sup>th</sup> century, considered by Friendly (2005) as “the Golden Age of statistical graphics”. Such profusion seems to have had an impact on journalism, with newspapers publishing the first pieces of information visualization.

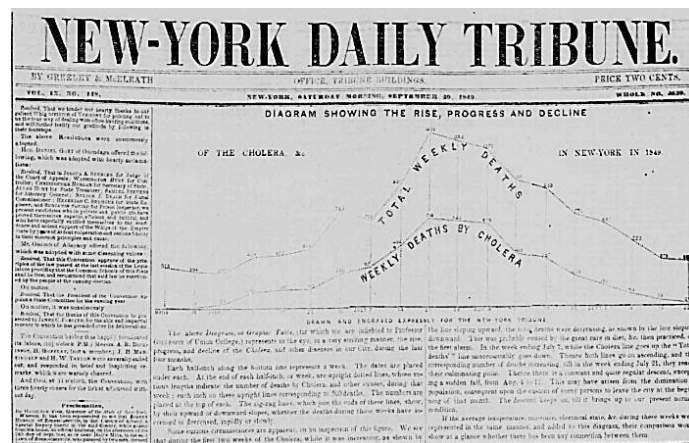
According to Simon Rogers (2013a, 2013b), *The Guardian* journalists have been working with sets of data – and trying to present them in interesting ways – since the British newspaper was first published, in 1821. Rogers shows, in the very first issue of the then *Manchester Guardian*, a long table of data (figure 6) listing every school in Manchester and Salford in that year, with the number of pupils that attended each one and average annual costs. “In 1821, it caused a sensation. Leaked to *The Guardian* by a credible source only identified as ‘NH’, it showed how official estimates of only 8,000 children receiving a free education were inaccurate – in fact, the total was nearer 25,000” (“The first *Guardian* data journalism: May 5, 1821,” 2011).

**Figure 6 - Manchester Guardian, 5th May 1821 (full version in [Appendix 1](#))**



Another influential example of early use of visual information in a newspaper is a chart published on the front page of The New York Tribune on the 29<sup>th</sup> September 1849 (figure 7), tracking the deaths in New York City from the cholera epidemic that summer. Attached to the infographic, a 300-word explanation of how to read it ([Figure 15](#), in Appendix 1) shows how unusual this kind of illustration was at that time.

**Figure 7** - Line chart showing the magnitude of cholera deaths in New York ("New-York Daily Tribune., September 29, 1849, Image 1," n.d.)



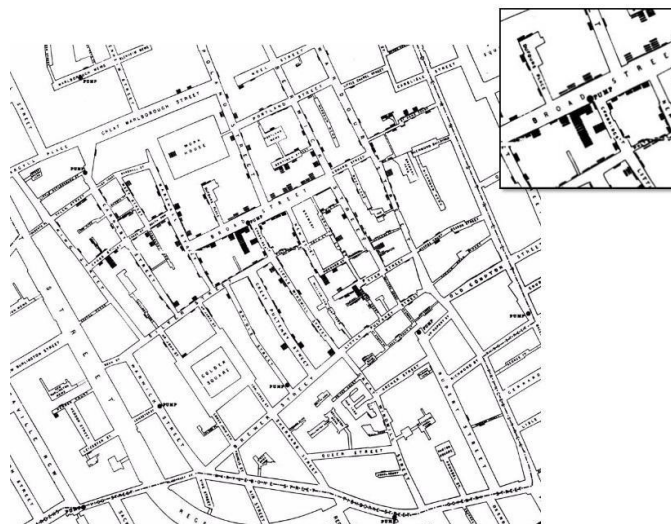
One of the most celebrated examples of visualization is also related to cholera deaths. John Snow's famous map of the disease outbreaks in London (figure 5) provided data journalists a model of how to work nowadays. In 1854, John Snow, an English

physician and expert in anesthesia and medical hygiene, plotted the five hundred deaths that occurred in a period of ten days in Soho neighborhood, central London. With his map, he demonstrated how the majority of cholera cases were located around a water pump located on the corner of Broad Street (now Broadwick Street) and Cambridge Street (now Lexington Street).

On proceeding to the spot, I found that nearly all of the deaths had taken place within a short distance of the pump. There were only ten deaths in houses situated decidedly nearer to another street pump. In five of these cases the families of the deceased persons informed me that they always sent to the pump in Broad Street, as they preferred the water to that of the pump which was nearer. In three other cases, the deceased were children who went to school near the pump in Broad Street. Two of them were known to drink the water: and the parents of the third think it probable that it did so. The other two deaths, beyond the district which this pump supplies, represent only the amount of mortality from cholera that was occurring before the irruption took place (Snow *apud* Tufte, 1997, p.6).

Snow solved the old mystery of cholera outbreaks because he had a good method: gathering and assessing evidence – data – besides a clear display of information. “Instead of plotting a time-series, which would simply report each day's bad news, Snow constructed a graphical display that provided direct and powerful testimony about a possible cause-effect relationship” (Tufte, 1997).

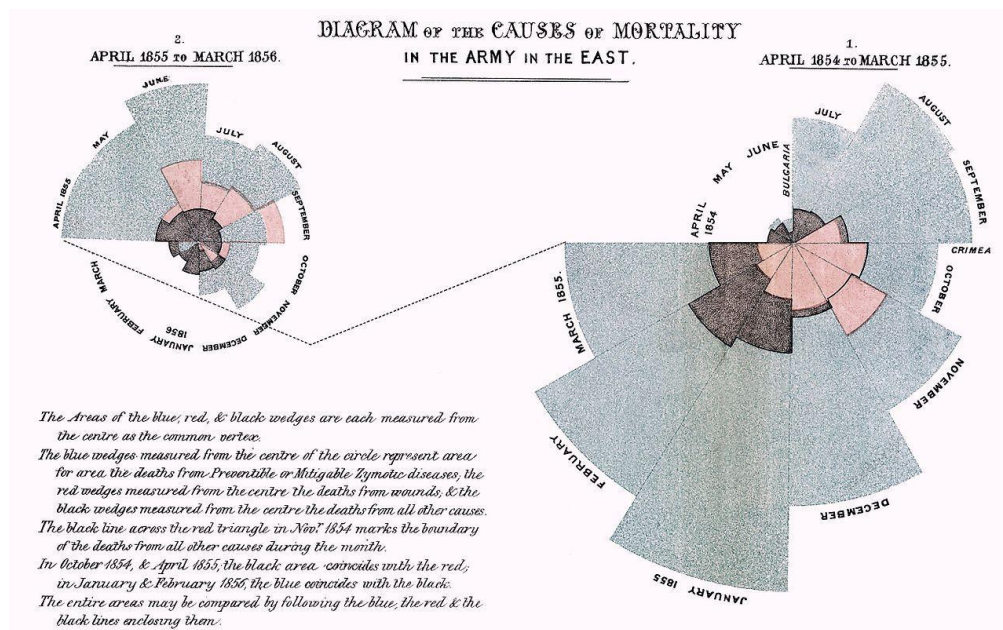
*Figure 8 - John Snow's cholera map of Soho, 1854*





Another striking visualization dates from 1858, after the Crimean war (1853-1856). The nurse, statistician, and reformer Florence Nightingale returned from the war and wrote down her *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army*, which she sent to Queen Victoria. It included a set of visual presentations of data, showing that the high number of deaths in Crimea were not directly linked to the war itself, and could be prevented. The Diagram of the Causes of Mortality in the Army in the East is similar to a pie chart – model developed by William Playfair in 1801 – to which Nightingale made an addition, by extending the size of the slices. Such drawing was a simple and clear way to explain complex statistics, and has become known as Nightingale's Rose Diagram (“Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army,” n.d.; Rogers, 2010).

Figure 9 - Florence Nightingale's Rose Diagram, 1858

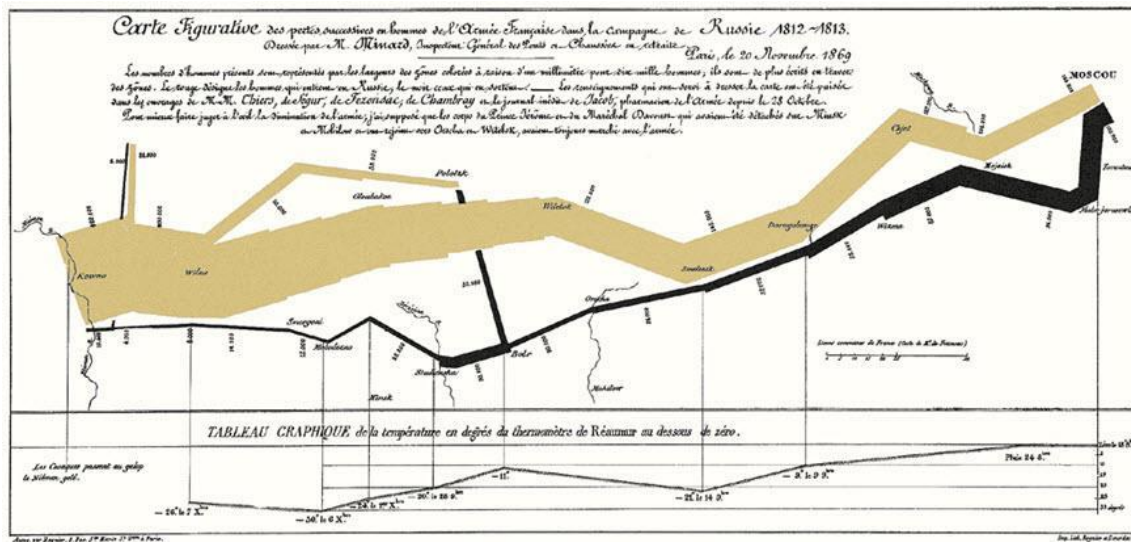


A few years later, an infographic by Charles Joseph Minard seemed “to defy the pen of the historian by its brutal eloquence” (E. J. Marcy *apud* Tufte, 1983, p.40). The French engineer successfully enhanced the power of time-series graphs, adding the spatial dimension to them. His classic chart drawn in 1861 displays the losses suffered in Napoleon's Russian campaign of 1812-1813. The width of the tape shows the size of the French army at each position. Napoleon's recoil from Moscow is illustrated by the



black lower band, which is linked to temperature and time scales. “Six variables are plotted: the size of the army, its location on a two-dimensional surface, the direction of the army's movement and temperature on various dates during the retreat from Moscow. It may well be the best statistical graphic ever drawn” (Tufte, 1983).

Figure 10 - Charles Minard's chart, 1861: losses of Napoleon's army in Russia, 1812-1813



In the first half of the 20<sup>th</sup> century, there were few graphical innovations. That is why Friendly (2005) called this period the ‘Dark Age’ of visualization. Alberto Cairo (2017b) argues that “a second Golden Age of visualization began in the 1960s and 1970s, with the work of people such as Jacques Bertin, author of *The Semiology of Graphics* (1967), and John W. Tukey, author of *Exploratory Data Analysis* (1977)”. For Cairo, we still are in this second Golden Age, and nowadays there are two classes of citizens: experts – statisticians, scientists, data journalists, business analysts, who are reasonably used to the visualization vocabulary – and the public, that is not.

This point made by Cairo brings an important discussion. In the big data era, with datasets being increasingly used for decision-making in all aspects of life, a person’s information visualization literacy is almost as important as reading and writing.

### 1.3.2 – Information visualization literacy

---

*It's not what you look at that matters, it's what you see.*

*Henry David Thoreau, American essayist, poet, philosopher, and historian*

---

Information visualization may be attractive and attention-grabbing, but if the population's average visual literacy is not considered, the story can get lost and end up alienating audiences. It seems that a large part of the population lacks skills and knowledge to make sense of visualizations.

Börner et al. (2016, p.3) define data visualization literacy as “the ability to make meaning from and interpret patterns, trends, and correlations in visual representations of data”. The authors conducted a study with 273 science museums visitors who were shown 5 out of 20 different charts, maps, graphs, and five network layouts, and found that, although participants were interested in the visualizations, they had significant limitations in identifying and understanding them.

The results (...) provide strong empirical evidence that a very high proportion of the studied population, both adult and youth, cannot name or interpret data visualizations beyond very basic reference systems, that is, they have rather low performance on key aspects of data visualization literacy (Börner et al., 2016).

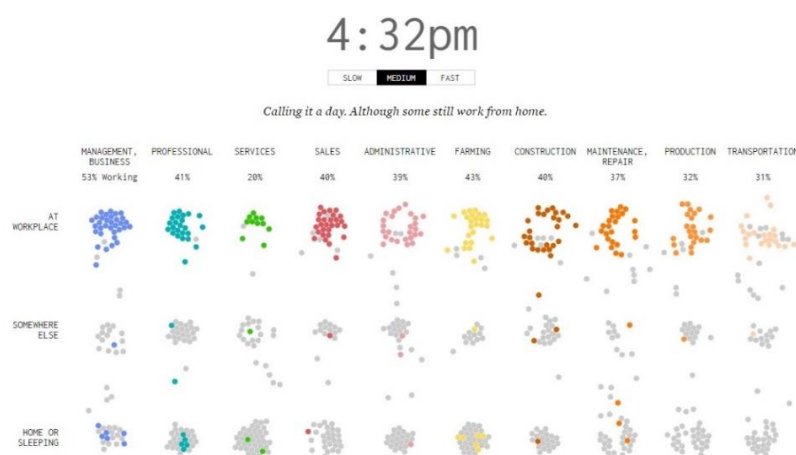
Similar outcomes were found by Maltese, Svetina, and Harsh. The researchers investigated the visualization literacy of 202 people with various levels of experience in science, engineering, technology, and mathematics. “Our results suggested that even those with advanced coursework in science and mathematics struggled with basic interpretation of common data visualizations” (Maltese et al., 2015).

It is important to highlight that the participants of the two studies were people with a predilection towards science – science museums visitors and professionals with degrees in related fields. This gives us insights on how low visualization literacy levels might be among general audiences.

However, because of the internet ubiquity, information visualization is no longer a practice restricted to science. As Gershon et al. (1998) predicted at the end of the last century, more diverse groups of users started using Infovis tools in a variety of contexts. Considering that it is a still growing movement, population's visual literacy tends to rise throughout the time. Furthermore, media outlets play a vital role in improving average information visualization literacy. The more data journalism uses visualizations to convey information, the more audiences get familiar with them.

Incidentally, the heavy use of information visualization out of the academy led Viégas and Wattenberg (2008) to coin a new term: *vernacular visualization*. The terminology was borrowed from the Tibor Kalman's idea of vernacular design, which is produced on the edges of the mainstream design. In this sense, vernacular visualizations are "visual representations of information created by and for non-expert users in contrast to those developed by expert system builders for specialized audiences" (Snyder, 2017, p.2097). Therefore, this category of Infovis has the potential to communicate complex ideas to a broader audience, while also working as a powerful educational tool. Figure 11 is an example of vernacular visualization. The animation by FlowingData<sup>15</sup> – based on data from 2011 to 2015 – shows a simulated day for employed Americans in different occupations. Each dot represents a person moving between home or sleeping, work, and somewhere else. Dots turn non-gray when respondents work and gray when they don't work (Yau, 2017).

**Figure 11** - *This is an American Workday, By Occupation* (Yau, 2017)



<sup>15</sup> <https://flowingdata.com/2017/05/17/american-workday/>

### **1.3.3 – Narrative visualization: when storytelling meets Infovis**

One of the first mentions of the marriage between storytelling – which can be understood as the act of unfolding a narrative – and information visualization was made in an article by Gershon and Page (2001). The authors argue that storytelling has the power to transform visualization in an effective and intuitive mean of revealing information and compare this process to watching a movie. The difference, according to the researchers, is that the information visualization environments are usually much more complicated than those typically shown in the traditional entertainment media.

The environments in which information visualization functions involve massive streams of information and data sources arriving in real time or from existing data and information sources (...). The problem in these environments is how to structure and present the information, so it is displayed efficiently, coherently, and economically, as well as what to include and what to leave out (the audience fills in the gaps). Moreover, presenting the information in a compelling and appealing way that enables it to be understood quickly is highly desirable. All this thinking, planning, decision-making, and data transformation and formatting means the resulting information visualization is more than a single image or animation clip; it's a kind of show business (Gershon and Page, 2001, p.33-34)

After nearly a decade, the topic grabbed the attention of a variety of areas, especially social communication industries. In the journalism context, the relationship between storytelling and information visualization has become so strong that it sprung a new denomination. Edward Segel and Jeffrey Heer (2010, p.1139) coined the term *narrative visualization* for those cases in which written stories are replaced by visualizations, emphasizing that “graphical techniques and interactivity can enforce various levels of structure and narrative flow”.

Lee et al. (2015) took a step forward and proposed three key aspects that characterize a visual data story (or a narrative visualization):

- Include a set of story pieces, which means specific facts backed up by data;
- these story pieces must be visualized – by annotations (labels, pointers, and text) or by narration – to support the main message;
- and must be presented in a meaningful order, or connected one to another.

According to these criteria, interactive visualizations that allow the viewer a completely free exploration are not considered narrative visualizations. It is necessary to include written explanations or annotations so that the audience can capture the author's message. Thereby, simple charts created from data can be turned into narrative visualizations by combining several of them and adding appropriate annotations and interactive resources to guide the viewer.

#### ***1.3.4 – Interactivity in information visualization***

The notion of interactivity has emerged with the creation of the World Wide Web. Since then, media researchers have exhaustively discussed its concept and role within online journalism. It is not the aim of this dissertation to deepen in the subject, as interactivity has become an ordinary feature of digital media. What is important to understand is how it adds value to the journalistic production (Deuze, 2001), more specifically, to information visualizations.

According to Jensen (2008, p.129) "interactivity is defined as a measure of a media's [*sic*] potential ability to let the user exert an influence on the content and/or form of the mediated communication". The author went further and divided interactivity into four categories:

- **Transmissional interactivity:** a measure of the media's potential ability to let the user choose from a continuous stream of information in a one-way media system without a return-channel and thus without the possibility of making requests;
- **Consultational interactivity:** a measure of the media's potential ability to let the user choose, by request, from an existing selection of pre-produced information in a two-way media system with a return-channel;
- **Conversational interactivity:** a measure of the media's potential ability to let the user produce and input his/her information in a two-way media system that is made available to other users, be it stored or in real time;
- **Registrational interactivity:** a measure of a media's potential ability to register information from and thereby also adapt and/or respond to a given user's needs and actions, whether they be the user's explicit choice of communication method or the system's built-in ability to automatically 'sense' and adapt.

Jensen (2008) elaborated such typology after analyzing information's production and distribution, as shown in table 1, where *center* can be understood as media outlet.

*Table 1 - Matrix of Interactivity (Jensen, 2008)*

	Information produced by center	Information produced by consumer
Distribution controlled by center	Transmissional	Conversational
Distribution controlled by consumer	Consultational	Registrational

Jensen's distribution control resembles a categorization by Segel and Heer (2010) that places most narrative visualizations along a spectrum between two extremes: **author-driven** (with a linear ordering of scenes, heavy messaging<sup>16</sup> and no interactivity) and **reader-driven** (with no prescribed ordering, no messaging and free interactivity). The authors highlight that the majority of them uses mixes of the two approaches, in three common structures:

- **Martini-glass:** begins with an author-driven path, using questions, observations, or text to introduce the visualization, and opens to a reader-driven interactive and explorative stage, resembling a martini glass design.
- **Interactive slideshow:** a typical slideshow format that incorporates interaction within each slide. This dissertation adopts the term **interactive sequence**, as it seems more appropriated to describe stories that presents a vertical unfolding.
- **Drill-down:** presents an overview and then allows the user to interact through tasks that reveal additional details and backstories.


The conversational and registrational categories were discarded for two reasons: first, to allow a comparison with other typologies, as both transmissional and consultational encounter parallels in taxonomies proposed by Infovis scholars (as seen

---

<sup>16</sup> Messaging in this case is the use of text by the author to guide the reader. It can be understood as annotations and explanations about the visual information.

in the table 2 and in the next paragraph); second, to simplify, since the majority of visualizations in data journalism does not allow the readers to input their own data.

*Table 2 - Interactivity taxonomies by author*

Segel and Heer	Jensen	van Wijk	Brehmer and Munzner	Barlow	Knafllic
Author-driven	Transmissional	Presentation	To present	Explanatory	Explanatory analysis
					
Reader-driven	Consultational	Exploration	To discover and analyze	Exploratory	Exploratory analysis

In an attempt to determine the value of visualization, van Wijk (2005) investigated technological and economic aspects, as well as the usage of Infovis. The researcher observed two main use cases: **exploration**, when users do not know what is in the data, and **presentation**, when a result is communicated to the user. Brehmer and Munzner (2013, emphasis added) stated that visualizations are driven by “either a need **to present** information or **to discover and analyze** new information”. Similarly, Barlow (2014) typified Infovis as **exploratory** – with an analytical value, created to help the user to figure out what is important within the data – and **explanatory** – detailed visualizations to reveal specific information to the user. Knafllic (2015) utilizes the same terms to address data analysis: **exploratory analysis** is the procedure of finding interesting information in datasets, while **explanatory analysis** is where the interesting information found via exploratory analysis is communicated to the intended audience.

Considering the tasks required to interact with an information visualization, one of the earliest and most notable approaches was developed by Shneiderman (1996, p.337) called Visual Information-Seeking Mantra – “overview first, zoom and filter, then details-on-demand”. It offers a valuable description of how the story pieces and/or annotations should be presented on screen to make the narrative visualizations most

effective for users. The author goes further and proposes seven tasks as a guideline for visualization creators:

- **Overview:** Gain an overview of the entire collection.
- **Zoom:** Zoom in on items of interest.
- **Filter:** Filter out uninteresting items.
- **Details-on-demand:** Select an item or group and get details when needed.
- **Relate:** View relationships among items.
- **History:** Keep a history of actions to support undo, replay, and progressive refinement.
- **Extract:** Allow extraction of sub-collections and of the query parameters.

Beyond a guidance for authoring information visualizations, it was the basis for a multitude of more extensive and complex taxonomies of interaction techniques, developed by several Infovis scholars to include new tasks eventually discovered (Yi et al., 2007; Boy et al., 2015; Figueiras, 2016; Young et al., 2017). As the scope of this dissertation is Infovis applied to data journalism, there is no need to employ thorough classifications, as the data journalism pieces rarely reach such refinement, whether because of the early stage of maturation of the field, the population's visualization literacy or other reasons that deserve more investigation. Therefore, some Shneiderman's categories were taken, others dismissed and two extra were added, seized out from and Boy et al. (2015):

- **Zoom:** show more or less detail.
- **Filter:** show information conditionally, by selecting from a drop-down menu, checking boxes or searching terms.
- **Details-on-demand:** show specific information, by hovering or clicking on elements.
- **Relate:** show relationships among items.
- **Reconfigure:** show a different arrangement, from a drop-down menu, or by clicking on elements.
- **Narrate:** show a different section, by clicking in a stepper-button or by *scrollytelling* (the mix of scrolling – moving text or graphics up, down, or across a screen – and storytelling).



Shneiderman's Mantra has also been the inspiration for many researchers developing interactive data visualization tools (Craft and Cairns, 2005), given his extensive experience designing information visualization software. Hereinafter, some tools are described in some detail.

### ***1.3.5 – Infovis tools: Tableau, D3.js, and Datawrapper***

Many authors have approached information visualization toolkits in their work. Most of the related literature dates from the middle 2000s and has been focused on developing sophisticated tools for building visualizations. "As the field of information visualization matures, the tools developed in our research laboratories are reaching users" (Shneiderman and Plaisant, 2006, p.2). And as these tools reach a broader public – outside academia – more user-friendly tools have been developed and researched.

Figueiras (2016, p.159) pointed out that there are Infovis tools "for every type of user, from experts to novices, and for every desired outcome, from traditional information visualization to narrative visualization". Those tools can differ in the programming language used (JavaScript, Python, R, etc.) and in type (library, framework, web application, toolkit, platform, etc.), and requires different levels of coding skills. Due to their popularity, three tools deserve attention: Tableau, D3.js, and Datawrapper.

**Tableau** is regarded as the grand master of information visualization software, with a customer base of over 50,000 accounts<sup>17</sup>, ranging from companies to private users. It offers an intuitive solution that comprises analysis and visualization of data, supporting a variety of charts, graphs, maps, etc. Some have called Tableau "the Excel of the 21<sup>st</sup> century"<sup>18</sup>, meaning that the Infovis tool represents a breakthrough in the same way as the spreadsheet software did in the 20th century. Tableau's limitation is that some features are only available in paid versions.

**D3.js** is a very popular JavaScript library. It is free but requires JavaScript coding skills. Its goal is not being easy to use, but to be a sophisticated and powerful tool that allows users to be more creative in the way they visualize data. The library was designed

---

<sup>17</sup> <https://www.tableau.com/about/customers>

<sup>18</sup> Elijah Meeks, Senior Data Visualization Engineer at Netflix, mention this expression in <https://medium.com/visualizing-the-field/the-7-kinds-of-data-visualization-people-9964e80443a7>

not only for visualization but also its distribution, as it can be flawlessly integrated into websites. In addition, it has a great collection of examples available.

Tableau's strengths are D3's shortcomings, and inversely, D3 shines where Tableau does not. It's easy to see that Tableau was built for the BA [business analyst], and D3 was built for the developer. However – that doesn't mean the two tools are mutually exclusive! We've seen and heard of some success stories with BAs using Tableau to explore and prototype their data visualizations, and then recreating those charts in D3 to create websites with an app-like reporting suite. This is a great way to combine the strengths of both tools and achieve great results (Paige, 2016).

**Datawrapper** is a commercial open source project, created by and for journalists “to bridge the gap between cool JavaScript visualization libraries and daily use in the newsroom” (Lorenz, n.d.). In late 2009, Mirko Lorenz<sup>19</sup>, a journalist who switched into digital project development, started working on a software that could simplify data visualization. ABZV, a German institution for journalism training, provided the fund and, after some trial and error process, Lorenz partnered with Nicolas Kayser-Bril<sup>20</sup> – one of the first to practice data journalism in Europe – to launch a beta version of Datawrapper, in early 2012. It reached the mark of 5000 users in the first three days. Gregor Aisch coded the 1.0 version, launched in November 2012 with more than a million visits in total. Nowadays, Datawrapper has a global user base of about 50.000 registered users, most of them professional journalists or communicators from other fields. The tool won the African News Innovation Challenge 2012 and the title "best start-up for news" at the Global Editors Summit in Paris, 2013. Datawrapper represents, thus, an important accomplishment, as it approximates journalists to data journalism's final stage, usually delegated to design and/or programming departments.

#### **1.4 – Data stories production process: roles and skills**

Several investigations have addressed what competences and skills professionals need to build stories based on datasets. A pioneering study conducted by Zanchelli and

---

<sup>19</sup> <http://www.mirkolorenz.com>

<sup>20</sup> <http://blog.nkb.fr/>

Crucianelli (2013) analyzed media outlets with well-established data journalism teams: four American (Washington Post, New York Times, USA Today e Chicago Tribune), one British (The Guardian) and one Brazilian (Estadão). The result was a report that identified four essential aspects for data journalism teams to work effectively in newsrooms:

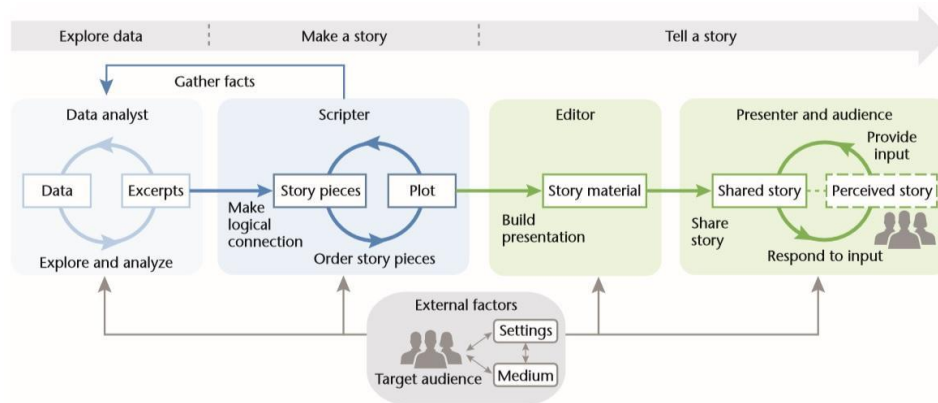
- Locating the team close to the news desk: it generates and improves collaboration between developers, designers, editors, and reporters.
- Encouraging all the team members to brainstorm data story ideas: as developers and journalists often have specialized skill sets, it's important to bring them together to increase the possibility of identifying unique and significant angles.
- Hiring professionals who bridge the skills gap: journalists who have some skills in data mining, development, programming, and design, as well as developers and programmers with notions of journalism and able to come up with stories suggestions based on data.
- Producing stories that reveal the meaning of that dataset and show its importance to the audience: choosing topics that affect the lives of news consumers drives web traffic, highlighting the importance of data journalism teams in the newsroom.

To delineate the workflow behind data stories, Lee et al. (2015, p.86) proposed a model called Visual Data Stories Process (figure 12), which “summarizes the main roles and activities that visualization storytellers engage in as they turn raw data into a visually shared story, along with the types of artifacts that result from these activities”. The model has three phases illustrated in a linear sequence, but they not necessarily happen in this order: exploring and analyzing data; making a story in an interesting and compelling way by organizing the data, establishing connections, formulating a message and creating the narrative; and telling the story, by building the final presentation, sharing it, and receiving the feedback.

Thus, Lee et al. (2015) identify four different roles in the process of turning data into a narrative visualization: the data analyst explores and analyzes the data; a scripter builds the storyline (plot) using the chosen excerpts; the editor prepares the story material with the chosen medium; the presenter delivers the story to the audience. The authors note that often one person plays multiple roles, but those roles can be

performed by different people. For instance, a company can hire statisticians to explore and analyze the data, journalists to build a plot, graphic designers to create infographics and programmers to build sophisticated interactive visualizations.

**Figure 12** - Visual Data Storytelling Process (Lee et al., 2015)



Weber and Rall (2012) also tried to outline narrative visualizations' roles. They interviewed 19 experts in eight media companies (five in Germany, two in Switzerland and one in the USA), focusing on their education, competences, tools, tasks, and responsibilities. It is noteworthy that the American company whose professionals were consulted was The New York Times, which is considered one of the state-of-the-art publications in the field. The respondents were information graphic designers, graphic editors, art and design directors, programmers, and journalists. They work in TV, print and online journalism outlets producing interactive graphics. The researchers identified that, in German and Swiss' newsrooms, the traditional concept of roles remains almost unchanged: journalists are responsible for the content or the story, designers deliver the visual part, and programmers deal with the coding required.

In contrast, the interviewees of the New York Times stated: 'We are journalists'. In this statement, we recognize a paradigm shift that has occurred in the New York Times newsroom. What is new is that even programmers and designers see themselves as journalists; they belong to the journalistic team of the newsroom and define their task as a journalistic task (Weber and Rall, 2012, p.351).

The New York Times' mindset evidences the increasing convergence of the field. Weber and Rall (2012, p.354) consider that the ideal narrative visualization professional, compared to the traditional journalist, must be "artistically skilled journalists who can serve the multi-disciplinary requirements of interactive information graphics almost autonomously, even though those persons are hard to find".

As Susan Reilly (2017) reminds us, The New York Times has full-time employees trained in statistical methods, data mining, coding and computer graphic, which give them abilities to find the story behind the data and make it visible. However, the author points out that many small local media companies cannot afford those kinds of professionals and, thus, these organizations end up redefining the existing journalists' jobs. The results often are overwhelmed journalists and non-satisfactory visualizations that turn to be a waste of the professionals' and readers' time.

To understand the state of data journalism today and its challenges, the Google News Lab<sup>21</sup> and PolicyViz<sup>22</sup> conducted 56 in-depth qualitative interviews with newsroom decision makers, data visualization experts, data journalists and video journalists in the United States, United Kingdom, France, and Germany. A large-scale online survey of over 900 companies in those four countries was also carried out. The result was a detailed report (Rogers et al., 2017) that gives us a good picture of the field: 42% of journalists responded they use data regularly to tell stories; 51% of news organizations have a dedicated data journalist on staff; and 53% of the professionals said that cleaning, processing and analyzing data are specialized skills that require extensive training or, in other words, it is not something that all journalists are able to learn easily.

Before this scenario, it can be said that there is a need for new journalism academic programs with interdisciplinary curricula, combining design, visualization, programming, and statistics. Some universities have already set up courses in data-driven journalism, but traditionally the academy has tended to respond slowly to developments in the industry. Not surprisingly, countries with a more developed data journalism market are the ones where with more advanced education in the field.

---

<sup>21</sup> <http://g.co/newsrab>

<sup>22</sup> <https://policyviz.com/>

Berret and Phillips (2016) collected information of 113 journalism course programs in the United States, including Puerto Rico. They found that 59 programs offer data journalism classes, ranging from teaching students only to use spreadsheets to analyze data for a journalistic purpose to multiple classes in programming skills, creating news applications, or building data visualizations.

But those more advanced programs were rare. Of the 59 programs we identified that teach at least one data journalism class, 27 of the schools offered just one course, usually foundational. Fourteen journalism programs offered two classes. Just 18 of the 59 schools offered three or more classes (Berret and Phillips, 2016, p.13)

Another study, based on curricula research and interviews with educators, was carried by Splendore et al. (2016) to describe the state-of-the-art of the current data journalism education in some European countries. The authors considered courses offered by academic, vocational, professional, and civic organizations. They observed that Germany, the Netherlands and the United Kingdom widely provide data journalism education in academic and vocational curricula. Moreover, these countries have professional institutions – such as media companies, unions, and associations – that offer short-term and on-the-job training. On the other hand, in Poland and Italy, data journalism education is modest, characterized only by civic training.

Specialized groups of data analysts or data journalists, such as the teams working for the Guardian, Times, Telegraph, The New York Times, or the Dutch RTL News Facts, are not yet a common practice – not to mention fully integrated data journalism departments in newsrooms that might coordinate their work with informatics or designers. Most media companies see data journalism as a cost rather than an investment. (...) All this may constitute a disincentive for educational initiatives in the field of data journalism (Splendore et al., 2016, p.149).

Despite being scarce, the research on data journalism education reveals that, even in countries where the field is more developed, incorporating data related disciplines into the universities curricula is imperative. Conversely, such inclusion is urgent where the practice of data journalism is still incipient.

## 2 – DATA JOURNALISM IN BRAZIL AND PORTUGAL

The present research intends, broadly, to provide an overview of the current development of data journalism in Brazil and Portugal and, specifically, to verify how information visualization is being used in the context of data journalism in both countries. Finally, the aim is promoting a cross-national comparison.

### 2.1 – Methodology

This dissertation resorts to a Qualitative Comparative Analysis (QCA), “an analytic approach and set of research tools that combines detailed within-case analysis and formalized cross-case comparisons” (Legewie, 2013). For the analysis of each unit, the case study strategy was drawn upon. As Yin (1994) observes, the use of case studies is justified when the subject is little investigated, requiring deep investigation of a few cases in order to set observation categories or generate hypotheses for future research. Thus, QCA is used here as a layer on top of an intensive multiple-case study.

The cases – Brazilian and Portuguese media outlets and data stories produced by them – were chosen considering the relevance of the journalistic publication in a national scale and the practice of data journalism (even if incipient). In Brazil, were analyzed three online versions of traditional newspapers ([\*O Estado de S. Paulo\*](#), [\*Folha de S. Paulo\*](#), and [\*O Globo\*](#)), two online-only news publications ([\*G1\*](#) and [\*Nexo\*](#)) and one independent initiative ([\*Volt Data Lab\*](#)). In Portugal, were investigated three online versions of traditional newspapers ([\*Expresso\*](#), [\*Jornal de Notícias\*](#), and [\*Público\*](#)), one online-only news publication ([\*Observador\*](#)) and one independent initiative ([\*Hemicycle\*](#)).

*Folha de S. Paulo* (known as *Folha*), *O Globo*, and *Estado de S. Paulo* (known as *Estadão*) respond for the highest circulation in Brazil (table 3) and all of them have data journalism teams. *G1* is a news portal, launched in 2006, that provides journalism content from various Group Globo outlets – such as Globo and Globo News TV channels, CBN radios, *Época* magazine, amongst others – and also produces its own multimedia content. It has a data journalism team. *Nexo* is an electronic newspaper launched in 2015, considered part of the independent Brazilian media and advocate of the Slow Media paradigm (Cunha, 2017), proposing a reflective, analytic, and deep journalism.

*Volt Data Lab* is an independent data journalism agency focused in investigation, analysis and data visualization that started as a blog at the end of 2014.

**Table 3** - Circulation of leading Brazilian newspapers with data journalism production

NEWSPAPER	CIRCULATION AVERAGE (December 2017)		
	PRINT	DIGITAL	TOTAL
Folha de S. Paulo	121,007	164,327	285,334
O Globo	130,417	112,987	243,404
Estado de S.Paulo	114,527	88,745	203,272

Source: IVC – Instituto Verificador de Circulação (Brazilian Audit Bureau of Circulation)  
Elaboration: Poder360/Drive.

*Expresso*, *Jornal de Notícias*, and *Público* are the major Portuguese newspapers with some data journalism production (table 4). Thus, other newspapers with high circulation but no data journalism work were not considered. *Observador* is an electronic newspaper launched in 2014, independently from the mainstream media, seen as a pioneering model in Portugal. *Hemiciclo* is an independent initiative of two professionals – not related to journalism, neither to the National Assembly – with the aim to bring citizens closer to the parliament work. Constantly updated and using publicly available data, the website gives access to the voting record of the 230 deputies in the last three legislatures through information visualizations.

**Table 4** - Circulation of leading Portuguese newspapers with data journalism production

NEWSPAPER	PRINT AND DIGITAL CIRCULATION IN 2017	
	MONTHLY AVERAGE	TOTAL
Expresso	45,533	546,399
Jornal de Notícias	26,165	313,988
Público	16,676	200,117

Source: APCT – Associação Portuguesa para o Controlo de Tiragem e Circulação (Portuguese Association for Copies and Circulation Control)



The information gathering was carried out using three methodological approaches:

- *semi-structured interviews* with data journalism professionals from all surveyed media outlets. A pre-determined set of open questions allowed for further exploration of subjects or responses and discussions of issues that may not have considered initially;
- *qualitative evaluation* of data stories produced by the chosen media outlets, as the goal is gaining an in-depth understanding of underlying reasons and motivations through more flexible techniques. To do so, a taxonomy based on the related literature was proposed, as can be seen further on.
- *quantitative assessment* of [DDJBR](http://ddj.br/)<sup>23</sup>, a list created by Volt Data Lab with the best data journalism stories of Brazil, according to Brazilian journalists from various vehicles. The list has monthly editions and the first was published in March. When this dissertation was being written, March, April and May editions were available and, thus, analyzed. As there is no equivalent list in Portugal, and the Portuguese data journalism production is sparser, the quantitative assessment was carried out only with Brazilian stories. Hence, the cross-national comparison takes place only qualitatively.

### **2.1.1 – Choosing interviewees**

In Portugal, finding appropriate interviewees included consulting staff webpages to find employees with titles related to *data* or *digital*, and then contacting them via Facebook, email or even LinkedIn. In Brazil, the search was more straightforward. After a Google search about *Estadão Dados* – the data hub of Estado de S. Paulo, which is widely recognized as the Brazilian pioneer in data journalism – the name of the person in charge, Daniel Bramatti, stood out. Having found his profile on Facebook, it was easy to contact him. Important to mention that he promptly replied and, later, added the present researcher in a virtual community – a WhatsApp group called #ddjbr (data-driven journalism Brasil) – where most of data journalism professionals are reunited to discuss and share experiences, knowledge, and ask questions. All the contacts for

---

<sup>23</sup> <http://ddj.br/>

interviews were drawn from the group and, moreover, many insights taken from there were helpful on this dissertation's development.

The professionals from Brazilian outlets were interviewed via online video chat, except for Thiago Reis, from *G1*, who ended up responding the questions by email, due to company's internal rules. In Portugal, professionals from *Público* and *Expresso* were inquired in person; Leo Xavier, from *Observador*, preferred to talk via online audio conference, Luis Vargas, from *Hemiciclo*, choose to speak by telephone call, and Pedro Pimentel, from *Jornal de Notícias*, asked to send the responses via Facebook Messenger.

*Table 5 - Data journalism professionals surveyed*

<b>MEDIA OUTLET</b>	<b>PROFESSIONAL</b>	<b>POSITION</b>
<b>Folha</b>	Fábio Takahashi, journalist	head of the intelligence department, in charge of academic research and data journalism
<b>O Globo</b>	Leonardo Cazes, journalist	executive-editor of checking and data journalism department
<b>Estadão</b>	Daniel Bramatti, journalist	head of Estadão Dados
<b>G1</b>	Thiago Reis, journalist	head of data journalism department
<b>Nexo</b>	Rodolfo Almeida, journalist	Infographic designer
<b>Volt Data Lab</b>	Sérgio Spagnuolo, journalist	founder
<b>Expresso</b>	Sofia Miguel Rosa, designer Raquel Albuquerque, journalist	Infographic designer reporter
<b>Jornal de Notícias</b>	Pedro Pimentel, designer	head of art department
<b>Público</b>	Rita Marques Costa, journalist	reporter
<b>Observador</b>	Leo Xavier, designer	head of technology and design
<b>Hemiciclo</b>	Luis Vargas, industrial designer and programmer	founder

All conversations took place in the Portuguese language, between the 8<sup>th</sup> and 20<sup>th</sup> March 2018, and lasted an average of 40 minutes (apart from those sent by email). They were recorded and later transcribed, also in Portuguese (see [Appendix 2](#)). Only the excerpts used in the body of this dissertation were translated into English.

The interviewees were asked from 10 to 15 open questions about their experience and backgrounds, the trajectory that led them to data journalism, the organization of the data journalism team within the company including tasks and roles, the data stories' workflow, tools and technologies used, constraints they experience in their work and so forth. Interviews with Sérgio Spagnuolo, from *Volt Data Lab*, and Luis Vargas, from *Hemiciclo*, were freer than the others, as their initiatives are apart from the mainstream media and, consequently, some of the guidance questions do not apply to them. Sérgio and Luis' responses are approached in the [section 2.4.4](#). All the others were divided into general themes, as can be seen in the [section 2.4.1](#).

### **2.1.2 – Choosing data stories**

For the qualitative evaluation, two criteria were used to select data stories. The first set of stories was obtained by interviewees' suggestions<sup>24</sup>. As they are expert data journalism professionals, nobody would be better to choose remarkable examples. They were asked to pick a recent data story, with no rigid constraints of data, as each company has a different data journalism production. The aim, in this case, was to identify the resources deployed to produce a flagship story. A limitation of this criterion is that exemplary pieces are not produced regularly and do not depict the routine of data journalism professionals.

For this reason, another group of stories was defined for each country, by similarity. The idea was to choose stories with a central subject in common and, as far as possible, the same database as a starting point. Finding the Brazilian sample was fairly easy, as all the analyzed news companies produced pieces about the high wages within the Judiciary, based on datasets published by the National Justice Council (CNJ), or by

---

<sup>24</sup> The exception was the *Jornal de Notícias* story, that was chosen by this dissertation author, because, according to the interviewee, the newspaper does not have a proper data journalism practice. Thus, the questionnaire was partially responded, as can be seen further on in the [section about the interviews](#).

regional courts following a CNJ determination. The only exception was *G1*, because – as explained the head of data journalism department, Thiago Reis – the editorial board decision for the department is working on special and in-depth projects. The *G1* stories related to the CNJ data were traditional pieces, written by other departments, with no data processing or visualization. Thus, for this sample, *G1* was excluded.

In Portugal, the choice was troublesome. It is worth to describe the process because it reveals how incipient and irregular is the data journalism production in the country. The first step was searching for stories by journalist. Using the name of the respective interviewees<sup>25</sup> on *Público* and *Expresso*'s websites, it was noticeable that the former produces more data stories than the latter. Thus, for each data story by Rita Marques Costa, from *Público*, a search by theme was conducted on *Expresso*, *Jornal de Notícias*, and *Observador*'s webpages. For example, Rita produced a [data story about the profile of the Portuguese drivers](#)<sup>26</sup>, based on datasets published by Portugal Motor Club (Automóvel Club de Portugal – ACP). Using the search option of the other three newspapers pages, stories based on the same study by ACP were found. *Expresso*<sup>27</sup> and *Observador*<sup>28</sup> used the data to write traditional text-only articles. *Jornal de Notícias*<sup>29</sup> reproduced one paragraph of an article from a website from the same media group, with a link to the original.

Another attempt was based on [Rita's story on Portuguese purchase power](#)<sup>30</sup>, based on data from National Statistical Institute of Portugal (INE). In this case, *Observador* used the same INE's database, as well as data from the statistical office of the European Union (Eurostat) and the Organization for Economic Co-operation and Development (OECD), to build [a special story](#)<sup>31</sup> including some visualizations about the

---

<sup>25</sup> Leo Xavier, from *Observador*, said there is no specific journalist in charge of data stories. Thus, the search could only be carried out by theme.

<sup>26</sup> <https://www.publico.pt/2018/01/23/sociedade/noticia/quem-e-o-condutor-portugues-conduz-nove-mil-quilometros-por-ano-tem-carro-com-uma-decada-e-usao-diariamente-1800028>

<sup>27</sup> <http://expresso.sapo.pt/economia/2018-01-23-Estudo-revela-que-so-10-dos-condutores-usa-transportes-publicos#gs.L6D6AZ0>

<sup>28</sup> <https://observador.pt/2018/01/23/estudo-acp-so-10-dos-condutores-usa-transportes-publicos-e-quase-metade-fala-ao-telemovel/>

<sup>29</sup> <https://www.jn.pt/motor-24/interior/portugueses-tem-cada-vez-mais-carros-mas-sao-cada-vez-mais-velhos-9071778.html>

<sup>30</sup> <https://www.publico.pt/2018/01/29/local/noticia/em-dez-anos-o-que-mudou-no-poder-de-compra-dos-portugueses-1799102>

<sup>31</sup> <https://observador.pt/especiais/o-outro-lado-do-salario-minimo-nacional/>

national minimum wage, within a section called *Observador Essays*. [Expresso](#)<sup>32</sup> and [Jornal de Notícias](#)<sup>33</sup>, again, presented only written pieces basically reproduced from the INE's report. Similar situations were verified when searching for themes as demography, emigration, consumption, and education.

Ultimately, the choice was based on a subject that caught the world's attention: the forest fires in 2017, considered the most tragic of an annual repeated phenomenon. The four Portuguese media outlets produced stories with some kind of information visualization. However, despite the amount of related data from previous years publicly available, the theme was very little explored visually and statistically speaking.

### **2.1.3 – Proposed taxonomy**

The qualitative evaluation of data stories samples consisted of assessing aspects that play an important role in data journalism projects, particularly when it comes to visualization. Inspired by Veglis and Bratsas (2017), and integrating elements from Segel and Heer (2010), Shneiderman (1996) and Boy et al. (2015), a taxonomy was proposed considering the presence of *some kind of visualization* (figure 13). Therefore, stories based on databases but presented using *only text, numbers and tables* – such as the ones mentioned before – were not considered for evaluation, as they would stop at the first stage of the hierarchical classification.

Having any visual element (map, chart, table or diagram), the next step was observing if the visualization was *part of the story* – complementing the narrative – or if it was *itself the story* – with text occupying the supplemental role of guiding the reader (Veglis and Bratsas, 2017). In the last case, the story's structure can be *martini-glass*, *interactive sequence*, or *drill-down*, as proposed by Segel and Heer (2010).

The visualization, part or center of the story, can be *mostly static*, resembling a printed infographic, or *mostly dynamic*, being in turn divided into *mostly noninteractive* and *mostly interactive*. If the Infovis is *mostly static*, it can be *simple* (presents only one

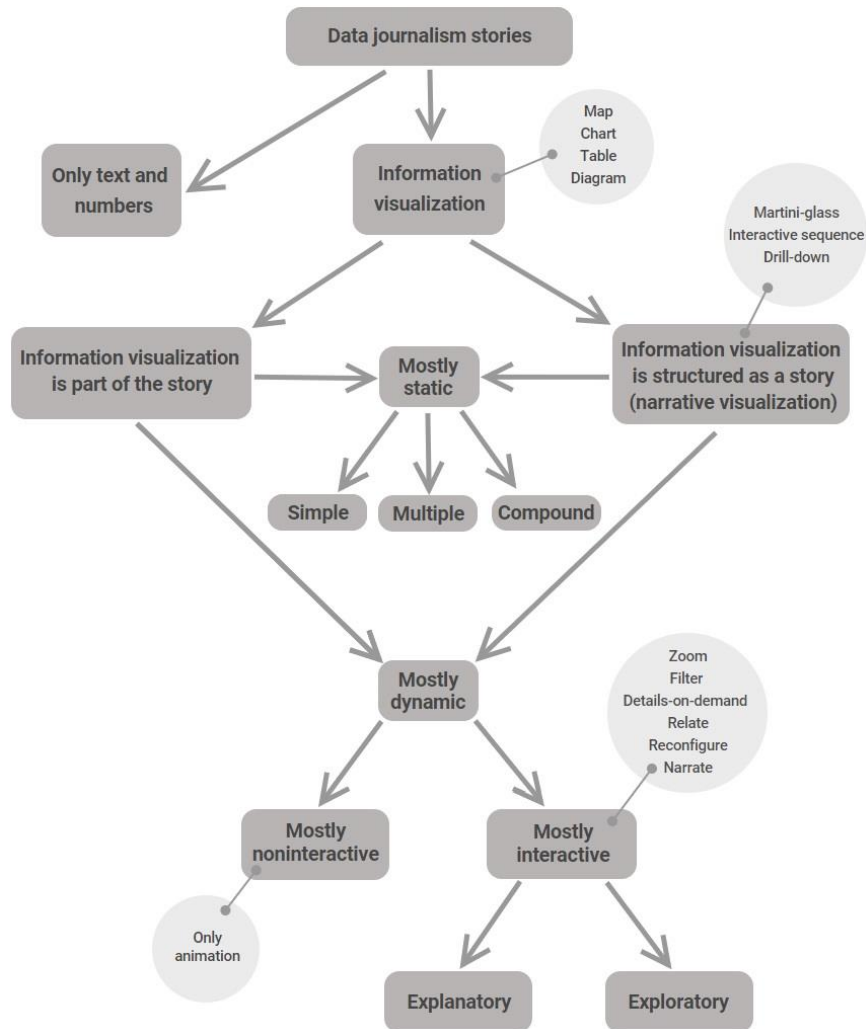
---

<sup>32</sup> [http://expresso.sapo.pt/economia/2017-11-13-Poder-de-compra-em-metade-dos-municipios-de-Lisboa-e-inferior-a-media-nacional#gs.srT\\_egg](http://expresso.sapo.pt/economia/2017-11-13-Poder-de-compra-em-metade-dos-municipios-de-Lisboa-e-inferior-a-media-nacional#gs.srT_egg)

<sup>33</sup> <https://www.jn.pt/economia/interior/poder-de-compra-em-portugal-sobe-para-773-face-a-media-da-uniao-europeia-8984727.html>

kind of graphic), *multiple* (presents more than one kind of graphic separately), or *compound* (presents different graphics merged into one or more compositions).

Figure 13 - Proposed hierarchical taxonomy



A *mostly dynamic noninteractive* story uses *animation*, which offers to the author the possibility of detailing information with no reader's action required. To be *mostly dynamic interactive*, the story needs to have interaction tasks, such as getting more information by hovering/clicking on elements (*details-on-demand*), clicking on a stepper-button to see the next part or scrollytelling (*narrate*), applying *zoom*, *filtering* content (by checking boxes, selecting an item from a menu or searching terms), finding relationships among items (*relate*) or *reconfiguring* the Infovis (Shneiderman, 1996; Boy et al., 2015). Being interactive, the story can be *explanatory*, if it is mostly author-driven,

or *exploratory*, if it is more reader-driven (van Wijk, 2005; Jensen, 2008; Segel and Heer, 2010; Brehmer and Munzner, 2013; Barlow, 2014; Knafllic, 2015).

This is not a rigid classification. Sometimes, a story can fall in-between two categories and have two balanced parts: one static and one dynamic, for example. In this case, the dynamic half could be considered stronger, and then the story would be mostly interactive. Thus, a level of subjectivity is necessary to overcome the deadlock.

## 2.2 – Brazilian historical context

The literature reveals some scarce initiatives of data-driven practices in journalism registered in Brazil in the 1990s, with the introduction of *computer-assisted reporting* (CAR) techniques. As Träsel (2013) reminds us, one of the pioneers in using this modality was José Roberto de Toledo, then a politics reporter working for Folha de S. Paulo. In 2002, Toledo founded, with other journalists, the Brazilian Association of Investigative Journalism (Abraji), which played a fundamental role in disseminating CAR concepts and techniques in Brazil, having trained more than four thousand journalists. For almost a decade, Abraji also strove for the right to access information.

[The Public Information Access law in Brazil](#)<sup>34</sup> was enacted in 2011 and implemented in 2012. In parallel, the popularization of tools for building, organizing, analyzing and visualizing databases fostered some data journalism initiatives. The first newspaper to recruit a team exclusively dedicated to data journalism was *Estado de São Paulo* (*Estadão*). In May 2012, the blog [Estadão Dados](#)<sup>35</sup> was created to be, according to its description, a data hub specialized in statistic-based stories and data visualization special projects. The first data-driven project – [Basômetro](#)<sup>36</sup> – was an interactive application that displayed, through information visualization, real-time congressional support for the federal government, based on the legislation voting records.

The pioneering spirit of *Estadão's* initiative was assigned to José Roberto de Toledo, due to his previous work with CAR techniques and interest for the field. As

---

<sup>34</sup> Law nº 12.527, 18 November 2011. Available in [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)

<sup>35</sup> <http://blog.estadaodados.com/>

<sup>36</sup> <http://estadaodados.com/basometro/>

relates the current editor-in-chief, Daniel Bramatti ([page xxxii](#) of Appendix 2), Toledo proposed to the directors' board the creation of a data journalism hub after attending an international conference by the Knight Center for Journalism in the Americas, from the University of Texas at Austin, with the best data journalism of the world. Bramatti explained that the principal characteristic observed by Toledo during the conference was the integration between journalists and programmers, what motivated him to suggest a similar composition for the about-to-be-created *Estadão Dados*.

Figure 14 - The first *Estadão Dados*' data-driven project: *Basômetro*



In August 2012, *Folha de S. Paulo* created the blog [FolhaSPDados](#)<sup>37</sup>, in partnership with the program Knight International Journalism Fellowships, from the International Center for Journalists. It was a temporary project – that lasted one year – created to analyze and visualize information, with infographics and maps, related to stories and articles published on the *Folha de S. Paulo* newspaper and website. The newly-created *Folha's* intelligence department is part of a recent effort of the

<sup>37</sup> <http://folhaspdados.blogfolha.uol.com.br/perfil/>



newspaper to reestablish the data team, with the recruitment of four data scientists coordinated by the journalist Fábio Takahashi ([page xxii](#) of Appendix 2). Differently from the older initiative – that was explicitly created to use interactive Infovis –, the current data hub was formed “to make headlines and great special stories, and not cool interactive infographics”, as explained Takahashi.

*O Globo* established a data journalism team between 2012 and 2013, at that time connected to the art and graphics department. According to the executive-editor of fact-checking and data journalism department, Leonardo Cazes ([page xxvi](#) of Appendix 2), after a recent restructuration, the team is currently apart from the art sector and linked to the newsroom, and its priority is fact-checking.

At the end of 2013, *G1* commissioned two reporters to work only with data, as mentioned the head of data journalism department, Thiago Reis ([page xxxix](#) of Appendix 2). According to him, that was the embryo of the current *G1* data journalism team, currently formed by five journalists.

Simultaneously with the creation of specific data journalism teams in the newsrooms, the field in Brazil gained an additional thrust with the arrival of the [School of Data](#)<sup>38</sup> – a network of data literacy practitioners composed of 13 organizations and 101 active individuals across Europe, Latin America, Asia and Africa. The global project is an initiative of the [Open Knowledge Foundation](#)<sup>39</sup>, a non-profit organization, active in more than 40 countries, focused on promoting open data’s value and helping civil society groups access and use data to take action on social problems.

In Brazil, [School of Data](#)<sup>40</sup>, promoted by the [Open Knowledge Brasil](#)<sup>41</sup>, organizes an annual data journalism conference, called Coda.Br<sup>42</sup>, in São Paulo. The first edition, on the 21<sup>st</sup> May 2016, assembled 120 participants and the second, on the 25<sup>th</sup> and 26<sup>th</sup> November 2017, congregated 250 attendees. The program has a hands-on approach

---

<sup>38</sup> <https://schoolofdata.org/>

<sup>39</sup> <https://okfn.org/>

<sup>40</sup> <https://escoladedados.org/>

<sup>41</sup> <https://br.okfn.org/>

<sup>42</sup> The event has a partnership with Google News Lab, Night Center for Journalism in the Americas, Brazilian Association of Investigative Journalism (Abraji), La Nación Data and Python Software Foundation.

and comprises workshops, boot camps, jam sessions, and talks. Although the initiative is still young, it raises the hope of a data journalism market's maturation in the country.

Such flourishing field grabbed the attention of Brazilian data journalism researchers (Träsel, 2013; Grandin, 2014; Lima, 2016; Vasconcellos et al., 2015; Mancini and Vasconcellos, 2016). All of them had at least *Estadão Dados* as its object of study, given its pioneering initiative. Träsel (2013) conducted an ethnographic evaluation of *Estadão Dados* professionals, to investigate their values, beliefs, and behavior. Grandin (2014) focused on the economic value created by the data journalism practice, also analyzing *Estadão Dados*. The other scholars presented a wider discussion, having investigated Brazilian mainstream media market in general.

Lima (2016) found out piecemeal initiatives of data journalism and observed that, despite the integration of specialized teams, few Brazilian mainstream media outlets invested consistently in data journalism.

Folha de S. Paulo's and O Estado de S. Paulo's projects show part of this reality. They are presented as parallel projects (blogs) or as special approaches of a given subject. Other outlets, such as O Globo, Zero Hora, Época and Valor Econômico also have good journalistic narratives based on data, but, just the same, they are isolated points within a journalistic trajectory present in Brazil for decades (Lima, 2016, p.296-297, own translation).

Similar results were found by Mancini and Vasconcellos (2016), who analyzed 2.296 news pieces of 48 online newspapers for five months. The authors noticed that *Folha de S. Paulo*, *Estadão*, and *O Globo* present typical data journalism pieces, whereas the other news websites use data only as a support. They also observed the main skills needed to data journalism practice: investigative competences (proactive data search and treatment), interpretative competences (ability to tell the story and reveal relations between the data) and communicative competences (data visualizations as a way to facilitate the understanding of the story).

### 2.3 – Portuguese historical context

The very few studies about data journalism in Portugal reveal the embryonic state of the field in the country, which seems to be far behind Brazil. There is no clear information in the literature about the first data-driven journalism initiatives. Some authors who investigated Portuguese media in general, such as Faustino (2004) and Bastos (2010), related that the newsrooms' computerization started in the middle 1980s. According to them, the newspaper *Público*, founded in 1990, was then a reference of such modernization.

It seems that the first formal data-driven initiative was registered between 2012 and 2014: a computer-assisted reporting project called REACTION, under the UT Austin-Portugal program, supported by the Portuguese Foundation for Science and Technology (FCT), in partnership with *Público* newspaper. The journalist Raquel Albuquerque ([page liii](#) of Appendix 2), currently working at *Expresso* and one of the interviewees of this dissertation, was at that time the Research Fellow of this project. According to her, it consisted of using computer tools to interpret and summarize datasets, and to create dynamic and interactive journalistic content based on data.

At *Expresso*, data journalism was formally included into the newsroom when Raquel Albuquerque was hired, soon after her fellowship finished. Before that, other journalistic initiatives that comprised some data analysis and visualization may have occasionally happened, as ad-hoc actions.

*Jornal de Notícias* presents a few data-driven stories, but its board does not recognize the practice of data journalism within the company. The head of the art department, Pedro Pimentel ([page lix](#) of Appendix 1), after having consulted directors and editors, concluded that the infographics team is the one in charge of such stories.

Although *Observador* has no professionals in charge of data journalism, the outlet has worked with data analysis and visualization since its foundation, in 2014, according to the head of technology and design, Leo Xavier ([page lxxv](#) of Appendix 2).

It seems that the first researcher to approach data journalism in Portugal was Ana Martinho (2013), in her master's degree dissertation. Martinho conducted a survey with directors of nine media Portuguese outlets, including generalist, economic and

sports newspapers, radios, and TV channels. The author also analyzed data stories of 17 outlets and found out that most of them were presented in form of traditional articles: blocks of written text with auxiliary static infographics. Not a surprise, given the fact that most of the directors surveyed for her research indicated printed media as the most suitable medium for stories based on data. However, an interesting point can be highlighted: “Sport is an area where there seems to be a systematized concern with data analysis and those responsible for this type of publications have given quite coherent and revealing answers that reveal their knowledge on data journalism” (Martinho, 2013, p.41, own translation).

Similar results were found by Ilo Aguiar Alexandre (2014). His master’s thesis was based on an analysis of daily generalist Portuguese newspapers and interviews with its editors. In addition to the scarce occurrence of data stories, the author observed that respondents could not even identify specific characteristics or skills needed for the data journalism practice. The only outlet with a team dedicated to collecting and analyzing data was *Público*, because of the REACTION project. Even so, Alexandre found only six data stories published by the newspaper in 2013.

### **3 – FINDINGS**

This section details the results of the *semi-structured interviews* with data journalism professionals, the *qualitative evaluation* of data stories produced by the chosen media outlets, and the *quantitative assessment* of [DDJBR](#), a list created by Volt Data Lab with the best data journalism stories of Brazil, according to curators from several Brazilian vehicles. Such findings are an overview of the data journalism state-of-the-art in Brazil and Portugal and show how information visualization has been used in the context of data journalism in both countries.

#### **3.1 – Interviews**

The semi-structured interviews offered valuable information at both micro and macro levels: it was possible to understand the production routine within each company surveyed, as well as to have a wider grasp of the data journalism practice in Brazil and

Portugal. One exception was *Jornal de Notícias*. Pedro Pimentel, head of its art department, apologized for not having responded all questions, justifying that, after consulting the board and editors, the conclusion was that the newspaper does not practice proper data journalism. Therefore, only three questions were answered, with information related to the infographics section.

An interesting finding of the interview with *Observador's* representative, Leo Xavier, was the company's innovative model – a kind of outsourcing or partnership with a spin-off<sup>43</sup> firm called *Frames* – for increasing the number of articles with graphics. Because it is unusual, *Frames* is approached in the same section as *Volt Data Lab* and *Hemiciclo* ([section 2.4.2](#)). Apart from those three initiatives, all the other subjects approached in the interviews were divided into seven general themes, as below.

### ***Initiative***

The officialization of data journalism teams or professionals has been always a company board's decision. However, in Brazilian newsrooms, such action was usually originated from a reporter's suggestion or after successful ad-hoc data journalism projects led by some members of the newsroom. A notable example is the launching of *Estadão Dados*. José Roberto de Toledo, then reporter, attended a data journalism event by the Knight Center for Journalism in the Americas, from the University of Texas at Austin, and observed an integration between journalists and programmers. Thus, Toledo proposed to the board of directors to establish a team with a similar structure. "It was promptly accepted because informally Toledo had done similar works before, and it was very good", highlighted Daniel Bramatti.

At *Folha de S. Paulo*, according to Fábio Takahashi, "there were individual initiatives but the data journalism team was structured by the newsroom head". Leonardo Cazes related that Fábio Vasconcellos – currently data journalism researcher and professor, and then a reporter at *O Globo* – had an important role in the first data team creation. "There was a support of an executive editor, Chico Amaral, a data journalism enthusiast, but the reporters alone have taken the initiative of training themselves, even before the implementation of a data journalism team", said Cazes. At

---

<sup>43</sup> A spin-off company is a new enterprise created from another firm, but is separated from the original organization.

G1, “reporters used to work with data autonomously (...) and the number of projects increased. Then, the direction, realizing its importance, created a nucleus addressed to data journalism in the newsroom”, narrated Thiago Reis.

At online-only publications, data journalism usually has emerged with the birth of the website. “Data journalism has always existed at *Nexo*, as an inherent part of the business model and editorial line”, said Rodolfo Almeida, journalist and infographic designer at *Nexo* ([page xliii](#) of Appendix 2). Although *Observador* does not have a data journalism team, Leo Xavier declared that there have been some special data projects since its launch, almost four years ago.

At the Portuguese companies analyzed, the official integration of data journalism seems to have emerged institutionally only at *Expresso* and *Público*, that have hired one professional each with the data journalist label. At *Jornal de Notícias* and *Observador*, there is a scarce data journalism production without this label. In practice, the daily news production absorbs the professionals recruited for this and data journalism ends up being relegated to the background. Raquel Albuquerque, from *Expresso*, said that she was hired because of her previous experience with data journalism at *Público*, but, in the end, what she does is journalism with data and not data journalism. According to her, except for few cases, “it is almost impossible to practice proper data journalism at a newsroom’s speed”. Rita Marques Costa, from *Público*, said that there is a lot of daily work which is not related to data and, as much as possible, she tries to “twist” the traditional stories with a bit of data work. Both journalists declared that the data journalism practice normally happens in parallel, when they have time to do it.

### ***Multidisciplinarity***

The journalists and programmers’ integration observed by José Roberto de Toledo in the United States reflects a fundamental characteristic of data journalism: it is a multidisciplinary practice. Whether in a team or in a professional – dedicated only to data journalism or not –, different abilities and backgrounds are needed, as it became evident both in the literature review and in the interviewees’ responses. Hence, the composition of the teams and/or the skills of the professional(s) might be complementary in order to encompass journalism, programming, design, and perhaps even some statistics.

Incidentally, the pioneer *Estadão* is currently going through a tough period, since the team is currently short-handed. It is a duo: Daniel Bramatti, the coordinator, and Cecília do Lago, both journalists with some notions of programming. “Because of the extra work of the next presidential elections [October 2018], we must hire a third professional, by May, who must be a programmer”, said Bramatti, reminding that the team once had five members, always multitasking professionals. Some of the professionals who were part of the team are: José Roberto de Toledo, as coordinator; Rodrigo Burgarelli, reporter and programmer; Lucas Maia, reporter with knowledge of statistics; Amanda Rossi, reporter with programming notions; Guilherme Jardim Duarte, programmer; and Diego Rabatone, a programmer hired exclusively for Basômetro, the first *Estadão*’s data-driven project, which nowadays is inactive. “It’s been six months without updating Basômetro, because the Parliament changed its website API [Application Program Interface] and also because some things were manual and now can be automated. This needs a programmer and we do not have one. In the beginning, we had three programmers who helped us to develop it. We were not able to hire someone to do this job yet, but I think it’s going to happen soon”, related Bramatti.

Some of the reasons of such difficult times are, according to the coordinator, the lack of funds and, ironically, *Estadão*’s pioneerism. This is because such well-prepared staff was gradually being hired by other companies, with higher salaries. However, being a pioneer also has its advantages, and one of them is that this shortage of resources forced the now diminished team to partner up with other newsroom’s sectors, and the result has been a naturalization of data journalism across the board.

Apart from *Estadão*, only *Folha* has had programmers/data scientists<sup>44</sup> to work exclusively on data stories. This is a recent undertaking, happening six years after *Estadão Dados* first multidisciplinary formation. *Folha*’s team is composed of one journalist – namely the head of the intelligence department, Fábio Takahashi – and three programmers/data scientists. Occasionally, journalists from other departments, for example politics, culture, economy, and so forth, are summoned to work along with this team, depending on the subject. Fábio himself said to have some coding abilities: “my programming skills are limited, but I think they are enough to talk to programmers, to

---

<sup>44</sup> According to Fábio Takahashi, they identify themselves as both programmers and data scientists.

bridge the gap between them and the newsroom and to ask the right questions for a huge database”. He also highlighted that one of the programmers – Daniel Mariani – has developed a journalistic nose for having worked at *Nexo* before, and that the newly recruited ones might acquire that notion along the way.

*O Globo* has a relatively large team of five members, but they are all journalists, except for the intern, who is a journalism student. It is important to highlight that the team is dedicated not only to data journalism, but also to fact-checking, which is the priority of the group. According to Leonardo Cazes, all the members have attended data journalism courses by the Brazilian Association of Investigative Journalism (Abraji). However, the lack of programming and design skills make the team dependent on the art and technology departments, which jeopardizes the data journalism production.

*G1* also has a journalists-only team. In addition to the coordinator Thiago Reis, there are four reporters, some of them with a basic programming knowledge. They rely on the art department, which has two programmers.

The journalist and infographic designer Rodolfo Almeida, from *Nexo*, presents himself the multidisciplinary that characterizes data journalism. In addition to his journalism and design training, he has a basic knowledge of statistics, programming, and data science. “I think it is an important combination, whether in the same person or in two people from the same team, of journalism and design skills”, affirmed Rodolfo, highlighting that this conjugation of skills is very helpful for the process of communicating the story behind the datasets. Aside from him, *Nexo*’s team has one data scientist and one journalism intern. “It is a department of data and research. *Nexo* has a model similar to the academic procedure, both in terms of sources and approach. Thus, the team generally comprises journalists and professionals with an academic career, such as Daniel Mariani, a programmer with a career in Biology who worked with us in the past and now works for *Folha*. Nowadays, our main production is information visualizations, but we also collaborate doing the research for the rest of the newsroom and special and/or interactive projects when they involve data”, explained Rodolfo.

In Portugal, as the companies have no data journalism teams, the professionals who produce data stories present multidisciplinary skills and/or work with other departments. At *Expresso*, both interviewees Raquel Albuquerque, reporter, and Sofia



Miguel Rosa, infographic designer, have acquired new abilities, such as coding basics, but not enough to overcome the absence of a programmer. Both professionals mentioned how important would be to have at least one developer, if not exclusive for data journalism, at least available for the newsroom's multimedia projects in general. An interesting point made by Sofia is that in the early 2000s when she began attending the annual Malofiej International Infographics Awards on behalf of *Expresso*, the infographic designers started to be integrated into the journalistic process and have gradually seen themselves as journalists. "From then on, we started considering us less as mere executors and more as coauthors of the journalistic information, which is not about a job position, but how you perceive your role", stated her. This perception matches with the mindset at New York Times, mentioned by Weber and Rall (2012), where programmers and designers see themselves as journalists.

Rita Costa declared that she took several online courses – visual storytelling, visual communication, spreadsheet software, python for journalists – in a two-year period, just before being hired by *Público*. In the newsroom, she affirmed that very often relies on the infographics department and sometimes on the web design sector.

At *Observador*, because there are no data journalists within the newsroom, the multidisciplinary of the designers and programmers coordinated by Leo Xavier becomes more evident, as he describes: "We are very autonomous, obviously always consulting the journalists, but they trust us. Their intervention occurs mostly in the suggestion. This is due to the many years we are on the team. In addition, I worked with other newspapers, so it is an area that I have always had an interest in".

### ***Location***

Data journalism's multidisciplinary leads to another significant issue: the geographic location of the team/professional, as pointed out Zanchelli and Crucianelli (2013), and most of the respondents confirmed.

At *O Globo*, the data journalism team was located within the art and graphics department until a recent restructuration, when data work was merged with fact-checking and the team has been positioned within a department called Radar, dedicated to producing hard-news to the online version of the newspaper. In Leonardo Cazes

words, “the goal is to contaminate Radar team with data and checking culture (...) and to improve the quality of data stories, providing them with contextualization”.

The proximity to the art department, previously verified in *O Globo*, is the current *G1*’s scenario. In its case, the closeness to the newsroom’s hard-news seems to be not so important, as its data stories are focused on deep investigations. Even though, this seems to be an exception among the surveyed media outlets that have data journalism teams. *Folha* has no separated room, what, according to Takahashi, allows for more conversations among newsroom’s members and a consequent dissemination of data culture. *Estadão* has a glazed room located inside the newsroom, but Daniel Bramatti highlights that the door is always open and everyone can come in at all times. “We work in partnership with Metropolis and Politics sections very often, and that is because the data work is spreading out the newsroom”, said Bramatti, adding that there are at least three journalists from those sections who practice data journalism<sup>45</sup>. *Nexo* has one space shared by all the employees, as related Rodolfo Almeida: “we are only 32 people, including administrative and maintenance staff, and we all work in the same office”.

In Portugal, none of the companies analyzed has data journalism teams. Thus, the professionals who deal with data do so inevitably within the newsroom. However, as they often depend on programmers and designers, the geographic position of the informatics and art departments influence their work. Raquel Albuquerque had experienced how important the team’s location is when she was the Research Fellow of the data-driven journalism project at *Público*: “I remember that the programmer used to work right beside us, we had lunch together and many times data stories ideas popped up at lunchtime”. Sofia Miguel Rosa, *Expresso*’s infographic designer, said that it is impossible to happen in *Expresso*, because the programming team is located in a different building, in another part of the city: “I normally meet them only when I have a technical problem that I cannot overcome on my own”.

---

<sup>45</sup> Actually, one of the *Estadão*’s data stories analyzed in this dissertation was carried out by the Metropolis section, with the support of *Estadão Dados*.

## **Workflow**

As Raquel Albuquerque described, the physical proximity between reporters and developers favors them to brainstorm data stories ideas. This is another key factor for a successful data journalism team, according to Zanchelli and Crucianelli (2013).

Most of the respondents mentioned that stories suggestions mainly come up from the data journalism professionals, either because they already have a structured or semi-structured database, or because they have a ‘trained nose’ to find relevant datasets. In this last case, which according to the responses represents the minority, data can be requested under open data legislation, scraped using specific tools, or collected by observation surveys, online forms or crowdsourcing. Then, the datasets are filtered and analyzed.

“At *Nexo*, some stories start with a database, for example from the Brazilian Institute of Geography and Statistics (IBGE) or the World Bank, or any other report. Others come from our curiosity about a theme and we go finding related data. Sometimes we do not find it, and the story does not go ahead. We do data scraping only as a last resort. The ideal setting is to work with structured or semi-structured databases”, described Rodolfo Almeida.

“In 60 to 70% of cases, the ideas come from us. We suggest and, if the editors accept, we produce and deliver”, explained Daniel Bramatti, from *Estadão*. The same occurs at *O Globo*, where most stories start within the data team. On the other hand, a minor part of the data journalism production come from editors’ demands, both at *Estadão* and *O Globo*. Interviewees from *Folha e G1* also cited the occurrence of the two flows abovementioned, without quantifying which one is more frequent.

In Portugal, the starting point of a data story follows those two dynamics as well. This when the daily hard-news routine gives space for it, as *Expresso*’s data journalist Raquel Albuquerque highlights: “if I finish the work I was designated to do earlier, then I can dedicate myself to search for interesting data, or have a look at a report that someone else told me about”.

Rita Marques Costa, from *Público*, affirmed that she might suggest more data stories than other journalists because of her sensibility to deal with data, but sometimes

editors come up with ideas. She also cited that ephemerides are good starts for data projects, such as Women's Day, International Day of Forests, and World Water Day. Similarly, Sofia Miguel Rosa, from *Expresso*, added periodic events suitable for data work, for instance, the Portuguese Schools Ranking and Elections.

At *Observador*, this is actually the start of a data story, as described by Leo Xavier: "When the newsroom has major demands, such as Elections, World Cup, Olympics or other sports events, we evaluate the possibility of doing the work in terms of deadlines and, if it is possible, we [designers and programmers] do everything, including data gathering, design, interactive visualization, and this is all mobile-first". Even though, Leo Xavier points out that this process happens two or three times per year.

Incidentally, all the Brazilian and Portuguese news outlets analyzed do not have a definite periodicity for data stories, apart from *Estadão* and *Nexo*. The first publishes data stories with static visualizations once a week, usually on Sundays or Mondays, and, parallelly, develops special in-depth projects, normally interactive, without a fixed deadline. The second releases three static narrative visualizations (visualizations structured as a data story) per week – on Mondays, Wednesdays, and Fridays – within a section called 'graphics'. According to Rodolfo Almeida, *Nexo* used to publish two data stories per week, with longer visualizations and more blocks of text, but the team realized that such format required a lot of hard work for little return. Then, the editorial decision was building right-to-the-point static graphics, which provide more time for the team to deliver three weekly visualizations. "When it comes to interactive visualizations, we try to produce a big interactive data story per month or, at least, every two months. The production can last more than a month. In this case, we rely on the technology department, because we need programmers who work with JavaScript", said Rodolfo.

### ***Technology***

On the subject brought by Rodolfo's quotation, JavaScript is the programming language used to build interactive visualizations. Hence, having professionals with this skill is crucial complete the last stage of data journalism's workflow: communicating the story. In some cases, the visualization can be a supplement of the text, but, as much as possible, it should act the as storyteller itself – a narrative visualization – with text playing a supplemental role.

Therefore, an overview of the technology used can offer a valuable clue about how advanced is the practice of data journalism in a media outlet. The pioneer *Estadão* has been using [D3.js](#), the most famous JavaScript library for visualizing data, since the beginning. As the team is currently incomplete, it has been relying on the infographics department, which also uses D3.js. For data processing and analysis, *Estadão Dados* resorts to Google Spreadsheets and SQLite (a lighter version of SQL, a standard language for storing, manipulating and retrieving data in databases). To build simple graphics, the team uses [Datawrapper](#) and to make visualizations with maps the choice is Carto, a paid software that provides georeferenced support.

*Nexo* also uses [D3.js](#) “only to build interactive visualizations, as the learning curve is steeper and requires more coding, but on the daily data stories we normally don’t use it”, said Rodolfo Almeida. For data processing and analysis, *Nexo* uses R, a free programming language and software environment for statistical computing and graphic, and sometimes Microsoft Excel. For static narrative visualizations, the choice is Illustrator, an Adobe’s vector graphics editor, that can be used in addition to QGIS – an open-source geographic information system application – if a map is needed. For smaller graphics, the team also uses RAWgraphs, an open source data visualization framework.

At *Folha*, data scientists basically use R. According to Fábio Takahashi, sometimes the team uses Microsoft Excel and SQL. He did not know which technology is used by the art department, and affirmed that the visual part is not so important: “We build visualizations few times [with the art department], because what the company expect is that we give a support for the newsroom to generate headlines or special stories, and not cool interactive infographics. This comes consequently.”

*G1* data journalism staff resorts to Microsoft Excel and Google Docs and Spreadsheets for analysis, Carto for visualizations with maps, and Timeline JS, a free and simple tool for telling stories in a timeline format. Thiago Reis mentioned the use of videos, animations and static graphics, but did not detailed the tools used. Nevertheless, he cited the proximity with the art department, that comprises two programmers. It can be drawn from this information that more complex tools, that require coding skills, might be also applied.

*O Globo* data journalists use Microsoft Excel “95% of the time”, according to Leonardo Cazes. As mentioned before, they rely on the art department for major data projects. To reduce such dependency, the team uses [Tableau](#) and [Datawrapper](#). “These tools allow us to circumvent the art staff, that don’t have enough time, because it covers demand from the whole newsroom”, declared Leonardo Cazes.

*Expresso*’s interviewees revealed the same reasoning. “We use [Tableau](#) to build some interactive visualizations, but it is limited and do not work properly on smartphones”, related Sofia, who also uses Illustrator and other Adobe tools, such as Photoshop, After Effects and Animator. Raquel referred using Microsoft Excel, Google Refine (for data cleaning and transformation to other formats) and some [Tableau](#).

Rita, from *Público*, uses [Tableau](#) as well, “even to make the analysis, because it’s cool to visualize data ongoing, instead of doing the graphics on Excel”. In addition, she uses ArcGIS, another tool for working with maps and geographic information. Rita also related that the web design department uses [D3.js](#) to build interactive stories and that the art department resorts to Illustrator.

At *Observador*, for the major data productions, the interactivity is provided by their own coding using JavaScript. Leo Xavier highlighted that journalists are encouraged to use Infogram, an easy to use web-based data visualization and infographics platform that requires no coding skills.

### ***Training***

A common point in both countries, as most of the interviewees related, is that the training and qualification were, so far, usually individual undertakings. The pioneer *Estadão Dados*, for example, is a result of the particular effort of José Roberto de Toledo. In addition, Daniel Bramatti revealed that journalists, and not only those from the data department, have learned data processing and analysis on their own, through courses held by the Brazilian Association of Investigative Journalism (Abraji) or online tutorials. According to Daniel Bramatti, “they realize that, without mastering it, it is not possible to do certain things”.

Nevertheless, this situation seems to be changing. Some of the respondents cited ongoing qualification initiatives provided directly or indirectly by the firms. Fábio

Takahashi said that *Folha de S. Paulo* is currently promoting a four-month free data journalism training for journalists from other departments and also from other companies. To qualify for the course, those professionals have undergone a selection process to verify their basic conceptions of programming and visual literacy. *O Globo* offered, in January and March 2018, a hands-on course on fact-checking and data journalism, aiming, “to spread the data and fact-checking culture across the newsroom”, as stated Leonardo Cazes, who admitted that there had never been a heavy investment in this field before. At *G1*, according to Thiago Reis, most of the training was carried out by the reporters themselves, mostly through free online courses, but the company stimulates journalists from other areas to get more familiar with data: “there have already been exchanges between reporters from other departments, and a course was developed for the branches spread throughout Brazil”.

In Portugal, besides the difficulty of practicing data journalism in parallel to the daily news production, both *Expresso* and *Público* are currently investing in their professionals. Raquel Albuquerque declared she is coursing an Information Visualization post-graduation course at Universidade Nova de Lisboa, on *Expresso*’s behalf. Rita Marques Costa said that *Público* offered her the same post-graduation, but she could not accept because of an ongoing master’s course, which she has started before being hired. Nonetheless, Rita affirmed that other two professionals at *Público* – a journalist and a designer – are enrolled in the same course. However, both Raquel and Rita related previous and current individual efforts to learn new valuable abilities on their own, such as basics of statistics and programming in Python.

### ***Challenges***

Most of the respondents mentioned two main difficulties for the data journalism practice: access to data and resource scarcity. Thiago Reis, from *G1*, well summarized the access issue. “The lack of transparency is still the rule. Often, even requests under the Public Information Access Law do not comply. Non-editable formats [when data scraping is required] are also common. Few institutions have repositories and historical databases available and/or easily accessible, which makes the work very difficult”. To that, Rodolfo Almeida, from *Nexo*, added the methodological low quality of some databases, what can induce errors. This is an issue also pointed by Rita Marques Costa,

from *Público*. According to her, in Portugal, open data initiatives are incipient – institutions do not publish their data and, moreover, do not collect enough data – and the information available very often presents errors.

When it comes to lack of resources, this seems to be the main constraint for the Portuguese companies' employees. Not having an exclusively dedicated data journalism team or professional clearly results in a sporadic and undermined practice. "I realized that the only way to do it is having an ongoing data project, with no deadline defined, in parallel with my daily work. I even think we can overcome the programmer absence using Tableau, but the reality is that there are a lot of limitations and we would like to do the work properly, in our manner", said Raquel Albuquerque, from *Expresso*. "That is when we need the programmer. It would be quite simple. It is just one more professional", completed Sofia Miguel Rosa, *Expresso's* infographic designer. Another point, added by Leo Xavier, from *Observador*, is the limited availability of multitasking professionals: "to have skills, you need time and money, so, in the end, it is everything related to lack of resources".

Another limitation cited was information visualization literacy, in line with the literature review. Daniel Bramatti, from *Estadão*, highlighted that "many people are not familiar with graphics, even the simple ones, and visualizations are usually elaborated, requiring a certain knowledge of how to interpret them". About that, Rodolfo Almeida, from *Nexo*, thinks it is one of the journalists and media's role: "an important part of our work is to educate the reader to learn interpreting graphics. There are still a lot of people with no visual literacy". However, still according to Rodolfo, *Nexo's* audience is more segmented and specialized than the public aimed by other media outlets, which allows the website to make a heavier use of visualizations with less amount of text. "There are people who follow us only because of graphics, so it is worth to approach subjects through visualizations rather than written text, because we notice that the most accessed stories are those with visual information", highlights him. This fact motivated the creation of a [monthly newsletter](https://www.nexojournal.com.br/noseixos/)<sup>46</sup> with the best infographics and visualizations, as well as articles explaining the process behind building them.

---

<sup>46</sup> <https://www.nexojournal.com.br/noseixos/>



*Table 6 - Snapshot of data journalism structure in the surveyed outlets*

MEDIA OUTLET	TEAM SIZE	PROFESSIONAL BACKGROUND	EXCLUSIVELY DEDICATED TEAM	MAIN TOOLS/ TECHNOLOGY
<b>FOLHA*</b>	5 members	1 journalist, 4 data scientists	yes	R, SQL, Excel
<b>O GLOBO</b>	6 members	5 journalists, 1 intern	no	Excel, Tableau, Datawrapper
<b>ESTADÃO*</b>	2 members	2 journalists	yes	Google Spreadsheet, SQLite, Datawrapper, Carto, D3.js
<b>G1</b>	5 members	5 journalists	yes	Google Docs, Google Spreadsheet, Carto, Timeline JS
<b>NEXO</b>	3 members	1 journalist/ infographic designer, 1 data scientist, 1 intern	yes	R, Illustrator, D3.js
<b>EXPRESSO</b>	1 member	1 journalist	no	Excel Tableau, Google, Refine
<b>JORNAL DE NOTÍCIAS</b>	Not applicable	Not applicable	Not applicable	Not informed
<b>PÚBLICO**</b>	1 member	1 journalist	no	Excel, Tableau, Arcgis
<b>OBSERVADOR***</b>	Not applicable	Not applicable	Not applicable	Infogram, JavaScript

\*At Folha and Estadão, other journalists (not members of the data journalism team) work on data stories sporadically, partnered with the team or not / \*\*At Público and Expresso, the only data journalist very often needs to rely on the infographics team to build more complex visualizations / \*\*\*Observador's data team is outsourced, but, sometimes, internal professionals build data based stories.

### **3.2 – Volt Data Lab, Hemiciclo, and Frames**

*Volt Data Lab* and *Hemiciclo* were deliberately included in the list of surveyed outlets because they are apart from the mainstream media. Thus, as independent initiatives, they give another perspective of the field, as well as broaden the data journalism panorama in both countries. Surprisingly, during this dissertation production, *Observador* has emerged with an innovative model to aggregate graphics to its articles by means of a spin-off company called *Frames*. Even though *Observador* is one of its

owners – hence, *Frames* is not quite an independent initiative – it was included in this section due to its peculiarity.

### ***Volt Data Lab***

Launched at the end of 2014 as a blog, *Volt Data Lab* is nowadays an independent data journalism agency. The journalist and founder Sérgio Spagnuolo ([page xlvii](#) of Appendix 2) explained that his interest in the field has emerged with the leaking of classified information from the National Security Agency by Edward Snowden, in 2013. “Data journalism and computational journalism were then becoming notorious and in Brazil those practices were incipient. I started learning programming using online tutorials. It was a trial-and-error process that required patience, but I insisted. In 2014, I worked as a freelancer during the World Cup and the presidential elections. Then, after these big events, I was a bit confused about what to do. If I went back to traditional newsrooms it would be difficult to work with data, because there was little opening to this kind of work. (...) So, I kept my occasional freelance jobs and started a blog to publish my ‘adventures’ in this field”.

Such ‘adventures’ led to more freelance jobs and partnership requests. The demand grew and revealed a business opportunity: Volt Data Lab has then assumed the format of a lean startup<sup>47</sup>. “I adopted this model because *Volt Data Lab* operates in a new communication segment in Brazil. The idea is hiring freelancers according to each job demand. Nowadays, I coordinate all projects and their methodological approaches and hire other people to do the gross part of programming”, related Sérgio.

The company exists for almost four years now, but, according to the founder, only in 2017 he started earning enough money to dedicate himself exclusively to the company. That was when Sérgio launched a selection process for a three-month fellowship. “Around 360 candidates appeared. I did not expect so many. Then, Renata Hirota, a journalist who is currently taking an undergraduate course in Statistics, has distinguished herself so much that she got the position and, after the three-month period of the fellowship, I invited her to work for Volt Data Lab with me. In addition,

---

<sup>47</sup> A methodology that can make the process of starting a company less risky and “favors experimentation over very elaborated planning, customer feedback over intuition, and iterative design over traditional ‘big design up front’ development” (Blank, 2013).

there is one more person, a programmer who deals with databases modelling and back end”. Nowadays, *Volt Data Lab* has as partners and/or clients some non-governmental organizations and media outlets, such as Piauí Magazine and The Intercept Brasil.

Additionally, the agency offers trainings on how to work with spreadsheets in an advanced level, data mining and data journalism foundations. “The academy is not being able to supply this market need for skilled professionals”, referred Spagnuolo. *Volt Data Lab* has just launched, in April 2018, a podcast called “Caixa-Preta” (Black-Box) which approaches the impact of technology on society, through interviews, stories, research and, of course, data. It was also a *Volt Data Lab*’s initiative the creation of a monthly list called [DDJBR](#), with data journalism projects developed by Brazilian media outlets. The list is curated by journalists from different vehicles and was object of the [quantitative assessment](#) carried for this dissertation. Another Sérgio’s endeavor was the aforementioned WhatsApp group #ddjbr (data-driven journalism Brasil). “There were 12 people when I created it last year. Now, there are over 130 professionals from many cities, news outlets, and organizations. There are solidarity, trust and collaboration between the members. If you have a question, just ask and someone will reply”.

To develop data projects, the team mainly uses R and SQL. Some projects were developed using Python and PHP, but those are exceptions. For static graphics, the option is R, and for the interactive visualizations, D3.js. Sérgio considers visualization is a very efficient way to communicate stories behind databases. “It is better than text and tables. (...) In addition, it is striking and helps the reader to explore the dataset. That is where interactivity comes in. When a visualization is interactive, it is itself the narrative, but the reader must have a knowledge to interpret it. (...) However, we develop a lot of static graphics, using R. To build an interactive visualization, the dataset must entail exploration. I do not use basic interactive resources, such hovering, because they require a lot of work for little results”.

When it comes to data journalism challenges in Brazil, Spagnuolo cited that the openness for this kind of content is still limited. He does not see access to data as a problem: “people complain a lot about the lack of transparency, but I think this situation is improving. Furthermore, I am in favor of journalists creating the own data through crowdsourcing or scraping instead of only depending on institutions”.

## *Hemiciclo*

The landing page displays a slogan that offers a good description of this independent project: “Democracy in real time”. *Hemiciclo* is a constantly-updated tool that uses publicly available data to show the voting record of each of the 230 elected members of the Portuguese National Assembly in the last three legislatures, through information visualizations. “It is a personal-initiative project, which aims to inform rather than express an opinion. The goal is, as much as possible, to be a neutral and analytical tool”, described Luis Vargas ([page lxix](#), Appendix 2) an industrial designer and autodidact programmer who has partnered with David Crisóstomo, an economist experienced in legislative process, to launch *Hemiciclo*. The website is similar to Basômetro, the first *Estadão Dados*’ data-driven project

The idea of developing *Hemiciclo* emerged to overcome a Portuguese Parliament’s breach. In the Declaration for Parliamentary Openness and Transparency, the National Assembly of Portugal has committed to ensuring an effective parliamentary monitoring and making available to the public a complete record of the individual votes of its members. However, this has not happened. “There is no tool that allows the effective scrutiny of the elected deputies’ voting. A solution would be someone from the civil society to carry it out. As a programmer, I found it a quite interesting challenge. I spoke with David Crisóstomo and we made it”, related Luis Vargas.

All the datasets used to build *Hemiciclo* are scraped automatically twice a day from the National Assembly website by a script developed by Luis. It is a robot that also organizes the information in the database. “The scraping was the most difficult part because the Parliament’s website is built in ASPX, an old technology. Thus, we needed a robot that could research, insert information in fields, store cookies and jump from page to page”. According to Luis, another challenge was structuring the information in a way that the website does not last a long time to access it. The solution was using a NoSQL database program called MongoDB. It is free and open-source. The visualizations of the semicircular benches<sup>48</sup> – with colored dots representing deputies and their parties – were built with [D3.js](#). The other elements were developed from scratch in JavaScript.

---

<sup>48</sup> The assessment of each element is in the section [Hemiciclo evaluation](#).

The webpage has no financial return, and its server and hosting costs around 15 euros per month. It requires a weekly maintenance, on Fridays, when voting usually takes place. “This is to adjust the data related to deputies who voted differently from their parties”, explained Luis, adding that he and David have ambitious plans for *Hemiciclo*. “We are including podcasts in partnership with a specialized group, and interviews with deputies. In addition, we will insert data from all legislatures since the 1990s and incorporate a section with visualizations of the parliamentary committees’ voting. Our goal is to apply for public funding so that we can dedicate ourselves exclusively to this project”. Luis Vargas does not rule out the possibility of a partnership with a media outlet as a way to capitalize the initiative.

### ***Frames***

When it comes to his relationship with *Observador*, Leo Xavier described himself as a two-in-one professional. He is the head of technology and design since the media outlet creation and, on the other hand, is also the founder of *Frames*. “Two years ago, we applied for the Digital News Innovation Fund<sup>49</sup>, that was seeking innovation projects in the field of online journalism. *Frames* has arisen from this project. Instead of setting it as an exclusive *Observador* undertaking, we decided to create a spin-off, a separated company, owned by *Observador* and myself.”

*Frames* product – that is, according to its founder, also called ‘frames’ – consists of small contextual graphics, sometimes with basic interactive resources (details-on-demand), incorporated in journalistic articles to improve their quality. “Our journalists pick relevant themes, search for data related and build simple graphics. For instance, if there were an earthquake in Mexico, they can produce a chart with the number of earthquakes in the region or even a graphical explanation of the Richter scale”. That way, a database of ‘frames’ has been generated and nowadays the archive has roughly 500 graphics. There is no defined periodicity to produce new ‘frames’, and the previous are constantly updated, so that the acquis is always up-to-date. The technology used at *Frames* is Node, a JavaScript environment.

---

<sup>49</sup> It is an European program that is part of the Google News Initiative, a partnership between Google and publishers in Europe to support high-quality journalism through technology and innovation.

*Observador* has started using ‘frames’ one year ago. Every time a journalist from its newsroom publishes a story, an algorithm created by *Frames* automatically analyzes it, classifies it, searches the archive for related graphics and, if it finds something that makes sense, publishes it. This is an automated process. A brief glance at some *Observador*’s stories with ‘frames’ shows that usually the inclusions of graphics are appropriately related to the articles’ content, even though sometimes such relations are not straightforward. For instance: in a story about investigated employees of the National Authority of Civil Protection, who might be somehow responsible for the biggest forest fire in Portugal in 2017, the chart included shows data of the National Authority of Civil Protection’s annual budget. Thus, the visual information is not directly linked to the article’s subject, but it is not totally detached.

If the *Observador*’s journalists find that the graphic does not suit the article, they can remove the ‘frame’. “This was deliberately projected to be automatic, because, according to our experience, if we rely on journalists to do the data work and to incorporate charts, maps, etc., we would have less than one percent of articles with graphics. Last year, when the only tool was Infogram, this rate was only 0.34%. With *Frames*’s system, in February 2018 we had 54% of articles with graphics. There is no newspaper in the world with this index”, narrated Leo, adding that such model is being exported for publishers from other countries.

### **3.3 – Data stories qualitative evaluation**

The samples for the in-depth evaluation were chosen considering two criteria: the first set of data stories was selected according to interviewees’ suggestions<sup>50</sup>, and the second was picked by subject similarity. Thus, the sampling comprises both flagship and simpler stories. A graphic by *Frames* was included within the sample of Portuguese stories related to the forest fires in 2017. *Hemiciclo* does not fit any criterion, as it has a single-issue approach. Hence, it was also evaluated separately. The assessment was carried out through the proposed taxonomy detailed in the [section 2.1.3](#).

---

<sup>50</sup> Except *Jornal de Notícias* story.

Figures with overviews of all stories are in the [Appendix 1](#), except *Estadão's* flagship story, whose interactive nature prevents any attempt of screening it. Below, the figures show frames of the stories, highlighted according to the point analyzed.

### ***Stories picked by the interviewees***

#### a) [Música muito popular brasileira](#) - *Folha de São Paulo*, 15<sup>th</sup> Dec 2017

This is a special story about what kind of music Brazilians listen to. After analyzing 134 billion plays between August 2014 and August 2017 on YouTube, *Folha's* data team traced the geography of fans of 340 national and international artists across Brazil. The dataset was scraped by an algorithm from 1,245 Brazilian cities (which represent 75% of Brazilian population) listed on the platform [artists.YouTube.com](#), that counts every time a song is played, whether in the artist's own video or the third-party's video.

The story was published online and on the printed newspaper. The printed version occupied eight pages of a special art and culture supplement called *Ilustrada*. This gives a dimension of the work behind the report – which involved, beyond the YouTube data, interviews and research conducted by eight journalists – and indicates, as can be seen in the story's overview represented in the [figure 17](#) (Appendix 1), that the visual treatment given to such rich content suits better the paper rather than the web. This fact is also evident in the few animations and interactive resources employed.

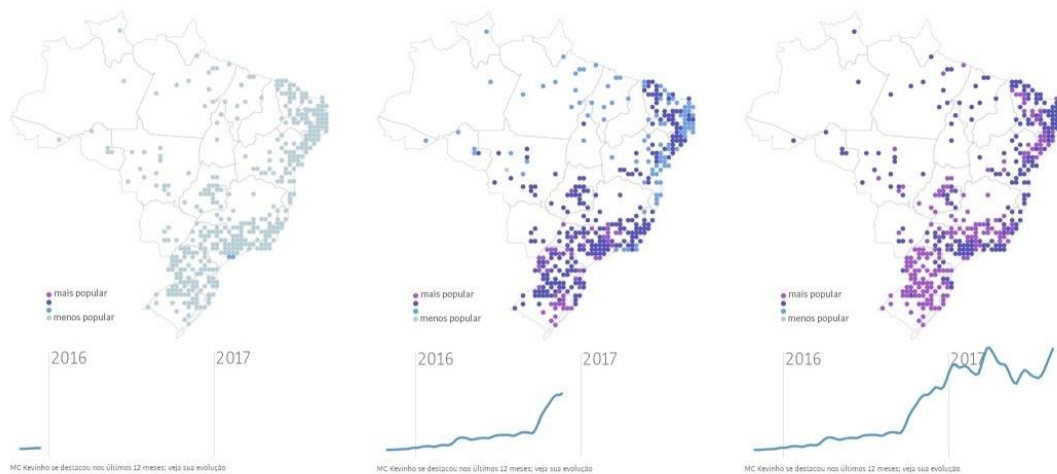
The whole feature is divided into several sections (one of them is the methodology of the data work) accessible by scrolling down the screen or through hyperlinks spread along the articles. Within those sections, there is a balanced quantity of text and graphics (mostly statics), as well as some videos and photos. Few animations are used. One is placed as illustration, positioned just next to the main headline, and consists of an automatic slideshow with 16 Brazil's maps, each one showing, in different colors, the geographic representations of various music styles.

**Figure 15** - The slideshow animation is built with the 16 maps below



Apart from that, three other animated graphic sets – each one formed by a map and a timeline – are used to show the evolution, in the last two years, of two artists' performances and of a hit by one of them.

**Figure 16** – The sequence shows the evolution of MC Kevinho is one of the three sets. The two other sets of map and line chart were not shown as they are very similar.

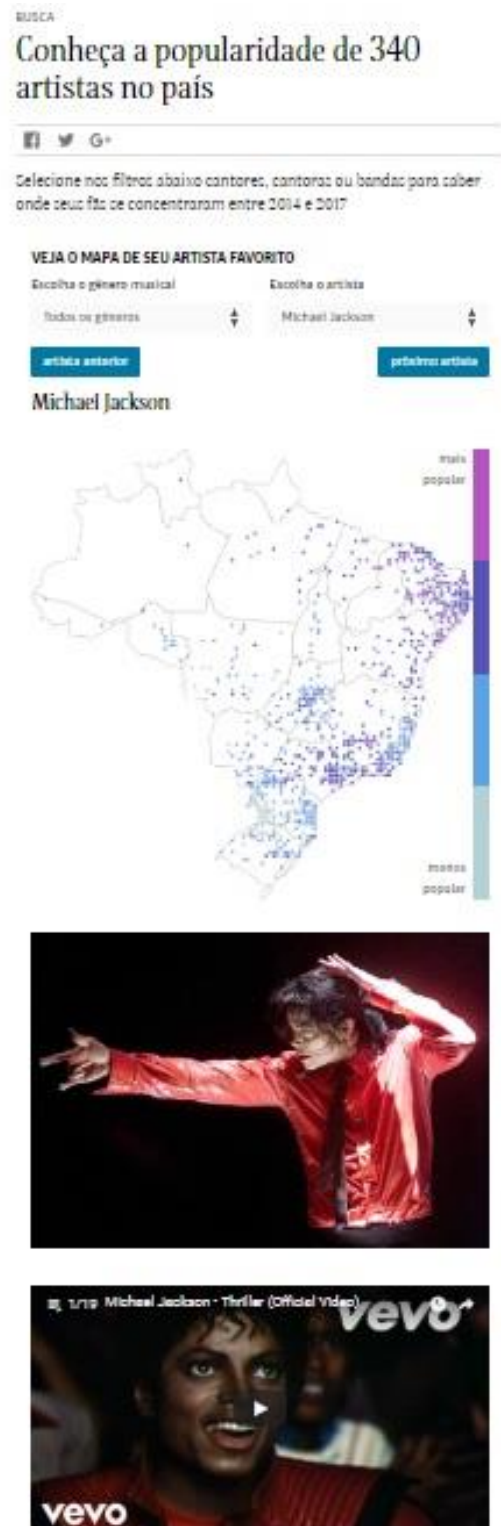


The only interactive part is a section called Busca (search), where the reader can see maps that show how popular artists are across the country. There are two drop-down menus: one *filters* artists by musical genre, and the other gives all the artists' names. Choosing from the first menu only restricts the list presented by the second, even though there are, in a later section, static maps with the various musical genres' distributions that could have been used interactively. There are also two buttons



(previous artist and next artist) for skipping along the list of artists. For each artist, beyond the map, there is also a picture and a YouTube video.

Figure 17 – The section "busca" is the only interactive part. The example below shows the popularity of Michael Jackson in Brazil, a picture of him and the YouTube videoclip of the hit "Thriller".



In addition to the visualizations build from the YouTube data (28 of them are static), there are a set of nine static graphics (five bar charts and four e pie charts) with data from the Music Consumer Insight Report 2017, by the International Federation of the Phonographic Industry (IFPI).

Summarizing, the visualizations are *part of the story*, and despite the use of interactivity in one section and some animations, they are *mostly static* and *multiple*.

*Table 7 - Summary of "Música muito popular brasileira", Folha*

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
37	4	no	yes	no	no	no	no

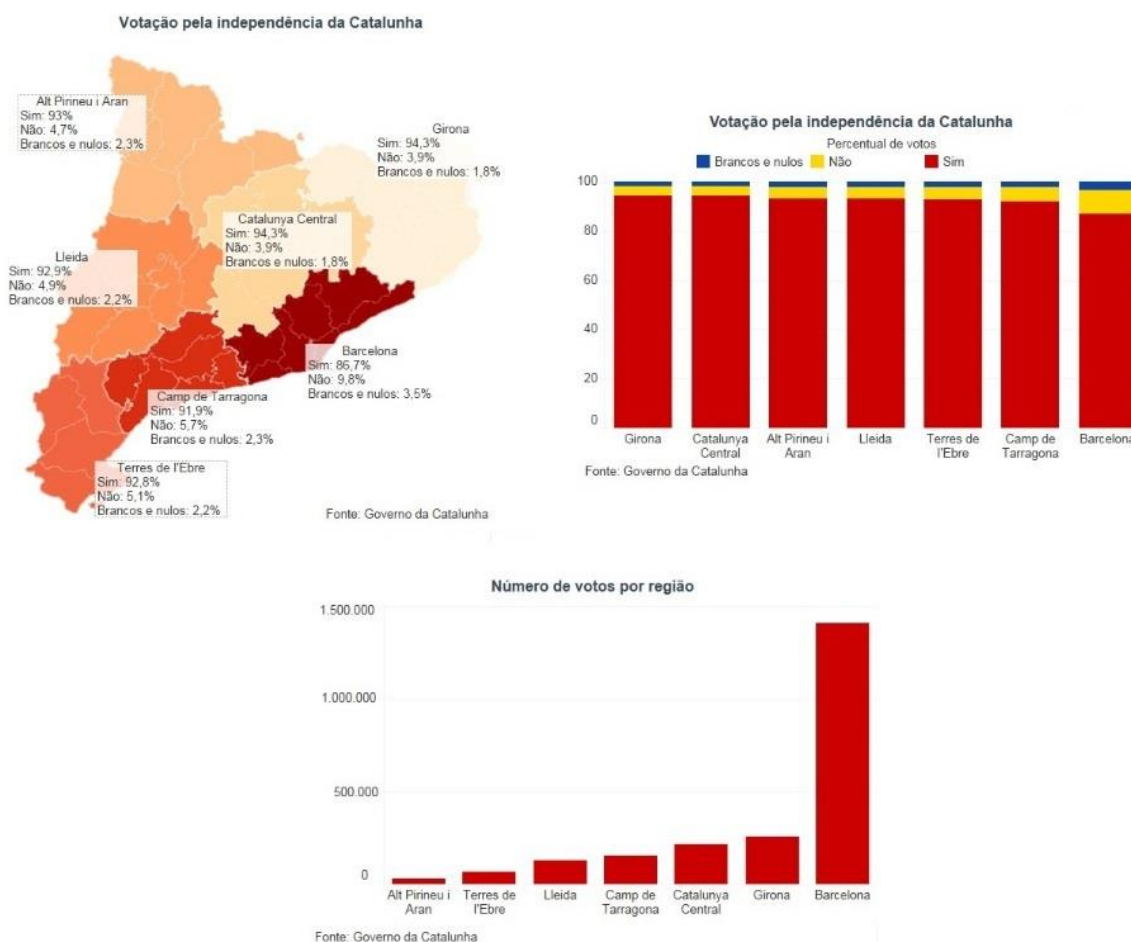
b) [Barcelona, capital catalã, teve maior proporção de votos contrários à separação](#)

- *O Globo*, 2<sup>nd</sup> Oct 2017

This data-driven story has three static graphics (one map and two bar charts) as illustrations for a written article about the results of the Catalonia independence referendum. Despite the size of the text being balanced in relation to the area occupied by the three graphics, the visualizations are used *as part of the story*, with a supplementary role. The visualizations are also *static* and *multiple*. The data source is the Catalan government, which indicates that the work of gathering, cleaning and analyzing were also reduced.

The layout of the story is in line with Leonardo Cazes' exposition about the productive routine at *O Globo*. The fact-checking has priority over the data work. In addition, as the data team has no programmers, the dependency on infographics and technology departments undermines the attempts of developing more sophisticated data stories and visualizations.

**Figure 18** - The story has only a static map and two bar charts



**Table 8** - Summary of "Barcelona, capital catalã, teve maior proporção de votos contrários à separação", by O Globo

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
3	-	no	no	no	no	no	no

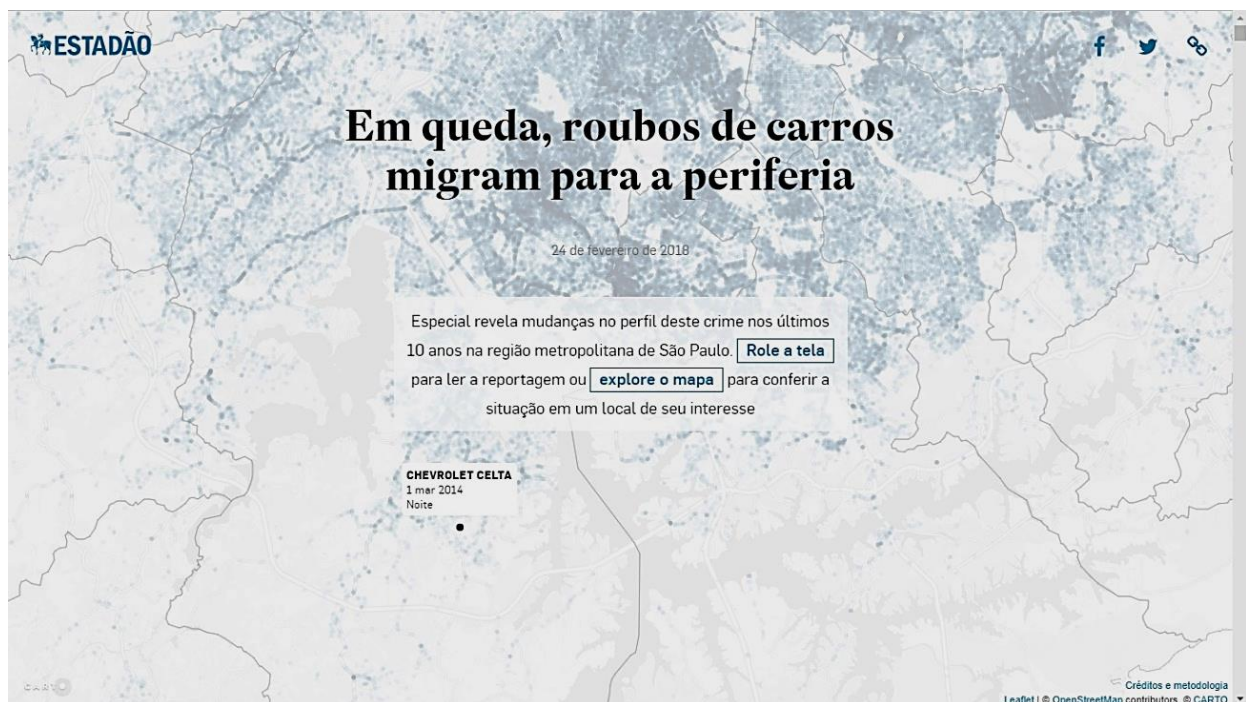
c) [Em queda, roubos de carros migram para a periferia](#) - *Estadão*, 24<sup>th</sup> Feb 2018

This *Estadão's* story – about the car thefts' migration to the suburbs of São Paulo – can be considered a model of narrative visualization. The information visualization is the center of the story, and some dynamic and interactive resources are employed within the narrative. According to the methodology, accessible through a link located at the bottom right corner of the screen, the source is the Transparency Portal of the São

Paulo State Public Security Department. The dataset comprises 365,000 registers of stolen cars in the 39 cities that form the metropolitan area of São Paulo, between 2008 and 2017. The professionals behind this story are not part of *Estadão Dados*. As Daniel Bramatti related, this is one of the cases in what the Metropolis section led the data work: “We gave a little help behind the scenes, but they did almost everything”.

The first view of the story presents a map with the hundreds of thousands of registers plotted on it. There is an animation that makes the data about the several stolen cars blink in different points of the map, representing the places where the thefts happened. The headline and a small explanatory text appear overlaid on the map. The explanation guides the reader to scroll down the screen in order to read the report or to click on a button to explore the map.

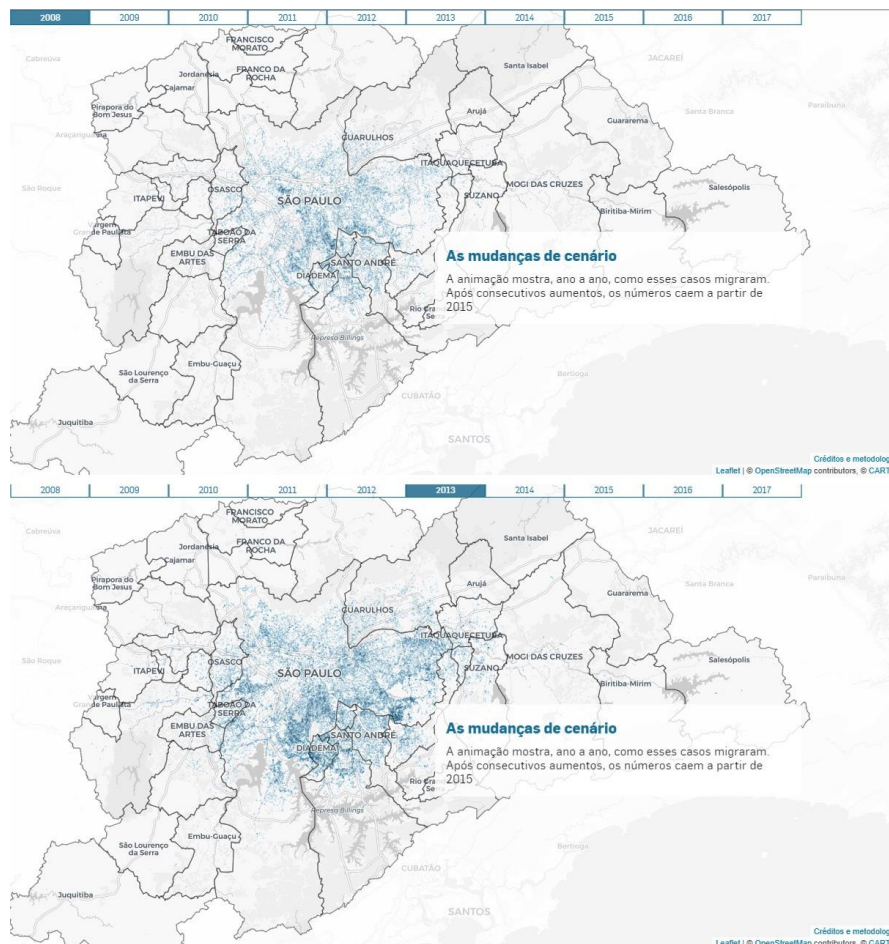
Figure 19 - First view of Estadão's narrative visualization



When scrolling down, small pieces of text transverse the screen as a layer on top of the dynamic map. Animations are used in different ways. The first scrolling, for example, is accompanied by a zoom on the map while a small piece of text gives an overview of the story. The second scrolling (figure 19) reveals an animated effect that

shows the migration of car thefts from the center to the suburbs of São Paulo city between the years 2008 and 2017. For each year, a different configuration of registers is shown. The speed of the animation gives the notion of how the prevalence of occurrences changed across the metropolitan area over the years. A three-line text complements the information.

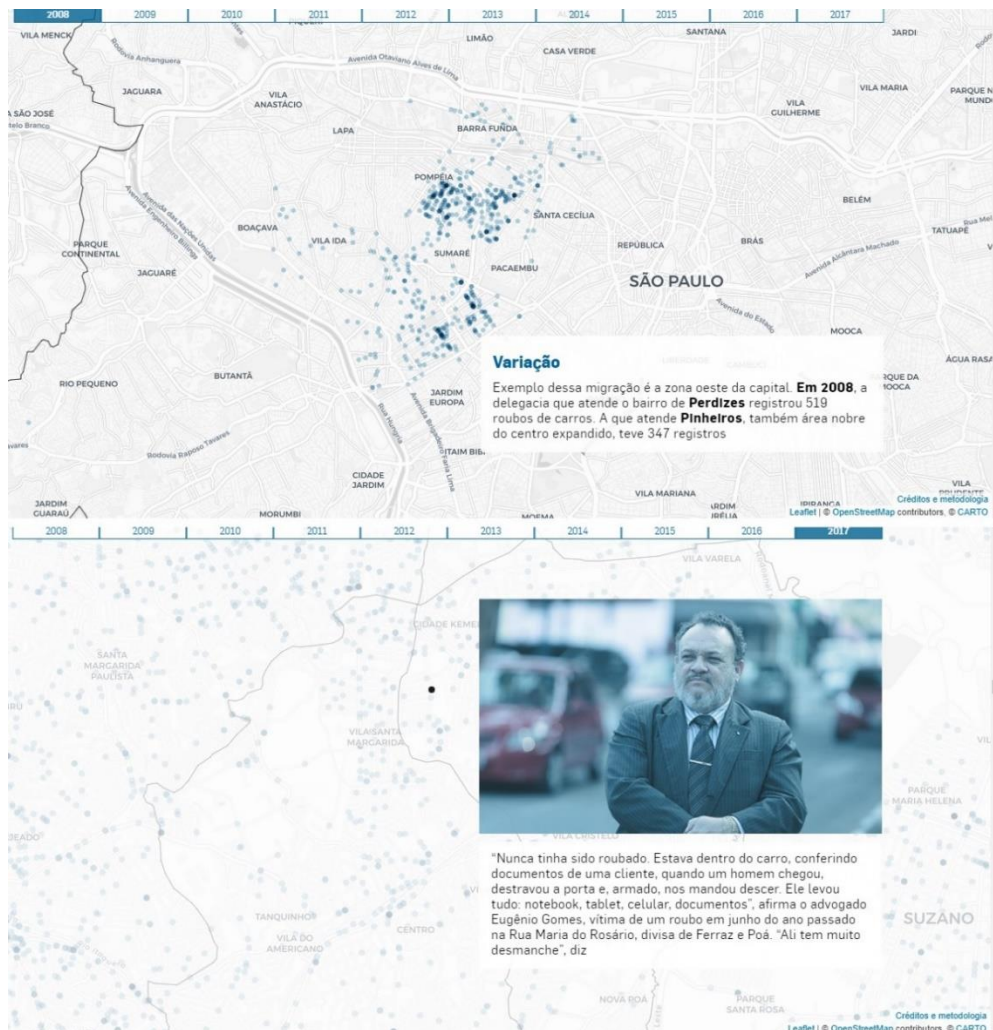
**Figure 20** - Animation shows the stolen cars in São Paulo between 2008 and 2017. Below, 2008 and 2013 maps are represented.



From then on, each screen has a different configuration, according to the year and/or information offered by the little chunks of text. For instance, the visualization presents details of some regions while the map is zoomed. In another screen, a photo of a stolen car owner and a quotation of him is showed over a map with a blinking point that represents the exact place where the theft took place.



**Figure 21** - The first screen shows a detail of the map with an explicative text. The second presents a photo and a quotation of a man who had the car stolen, and, in the background, the map with the exact point where the theft happened.



Altogether, the story comprises 42 screens, each one with a piece of text. Two of them have also a photo. The final screens added colors on the background map and line charts. Even a table is shown. Thus, besides having a considerable volume of text, the information is unveiled in a way – through scrollytelling and animations – that the text does not upstage the visualization.

The penultimate screen brings the instructions for the story's climax, driving the reader to continue scrolling once more to browse the map. The options are searching for a specific address and filtering the results by vehicle model, time, or period over the 10 years. The results can be shared on social networks.

**Table 9** - Summary of "Em queda, roubos de carros migram para a periferia", by Estadão

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	multiple	yes	yes	yes	no	no	yes

This is another example of narrative visualization. It is part of a special section on G1's website, called Monitor da Violência (Violence Monitor), that comprises several features on the subject. For this story specifically, the datasets from 2011 to 2016 were drawn from the Brazilian Forum of Public Security yearbooks. From 2017 on, the data collection and its processing has been carried by G1, and constantly updated. One journalist in each Brazilian member state oversees such work.

The first view of the story presents monthly data of the whole country – a rate per 100,000 inhabitants – about violent deaths (homicides, robbery followed by death and bodily injury followed by death), both by colors plotted on Brazil's map and in numbers, next to the map. The data can be *filtered* according to three drop-down menus: type of crime, month and Brazilian member state. It is also possible to choose the member state by clicking on it. For each member state, there is the number of crimes written next to map, two videos (with analysis from the state government and from an expert in public security) and a small bar chart, where each bar represents the number of crimes in a month of the current year. In addition, it is possible to compare the showed state with another, picking it from a lateral drop-down menu.

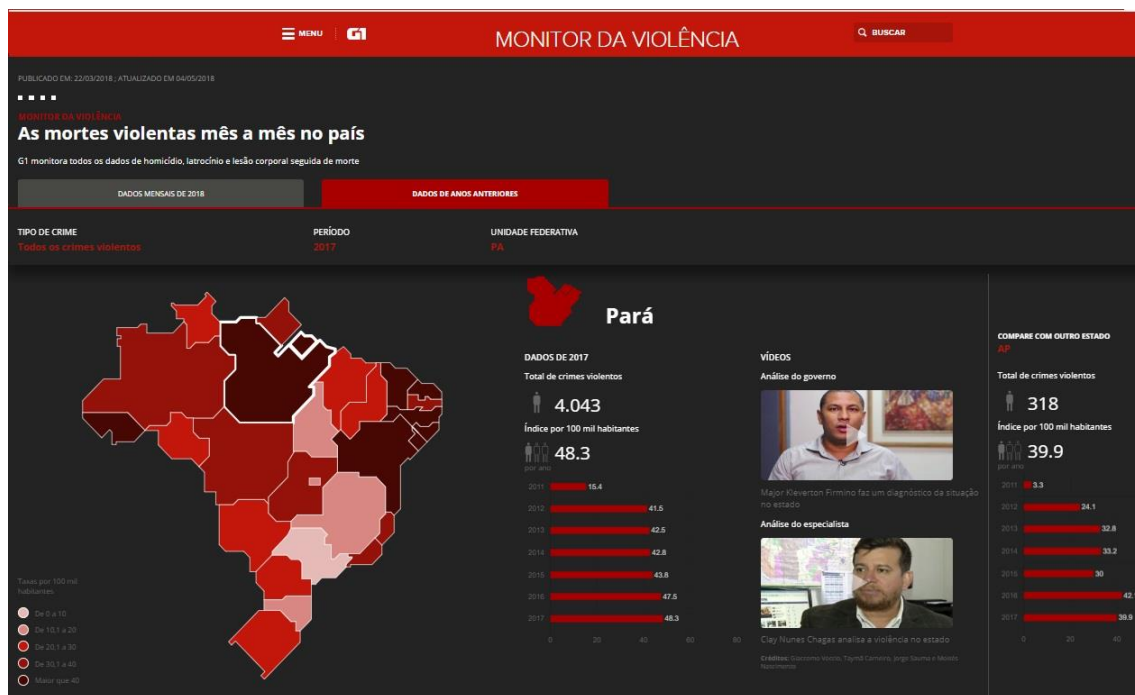
**Figure 23** - G1's story presents four drop-down menus and one tab that leads to an annual search



Furthermore, clicking on another tab, the reader can search data from previous years (2011 to 2016), by type of crime and Brazilian member state. In this case, there is no monthly filter. The comparison between states is also available.



Figure 24 - The annual data of violent deaths in Brazil



Summarizing, in this G1 story, the visualization is the *center of the story* and *mostly interactive exploratory*. Hence, it is a narrative visualization with a *drill-down structure* (Segel and Heer, 2010), because it presents an overview first and then allows the reader to access additional details and backstories.

Table 10 - Summary of "As mortes violentas mês a mês no país", by G1

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	no	no	yes	no	no	no

- e) [Como sua cidade gasta dinheiro? Descubra e compare com outras](#) - Nexo, 20<sup>th</sup> Sep 2016, constantly updated

The headline of this Nexo's story outlines the interactive tasks the reader must perform. It asks the user "how does your city spend money?" and requests the him/her to discover it and compare to other cities. This is intentional, according to Rodolfo

Almeida: “after analyzing the data, the first thing we do is the headline, which must be like an umbrella that covers all the information the reader can get from the story”.

A short three-paragraph text describes what is the municipal budget and how it is defined, reveals the data source (Center for Metropolitan Studies) and the most recent year of data collection (2014 or 2015, depending to the city), and explains how to read the graphics. Then, guides the reader to choose a city (*filter*) typing its name in a search box with auto-complete function. There is also the option of picking one more city in order to compare with the first municipality.

Figure 25 - Nexo's story first view, with a short introductory text and search boxes

**NEXO**

INTERATIVO

## Como sua cidade gasta dinheiro? Descubra e compare com outras

Daniel Mariani, Ibrahim Souza, Wellington Freitas e Rodolfo Almeida 20 Set 2016 (atualizado 03/Mai 15h44)

Com dados dos gastos dos últimos anos, veja como seu município optou por gastar seu orçamento comparando com outras cidades ou com as médias do seu Estado e do Brasil

CONTEÚDO PRODUZIDO EM PARCERIA COM O CENTRO DE ESTUDOS DA METRÓPOLE

Todos os anos, prefeitura e câmara legislativa definem juntas como será usada a receita de uma cidade. A elaboração do orçamento municipal se dá por meio de um projeto de lei que parte do executivo e pode ser aprovado e modificado pelos vereadores. É nesse documento que fica estabelecido se a cidade investirá mais em educação ou transporte, por exemplo.

Em parceria com o CEM (Centro de Estudos da Metrópole), o Nexo separou o gasto anual executado mais recente (2014 ou 2015, de acordo com dados disponíveis) dos municípios brasileiros por área de aplicação (saúde, transporte, etc).

Abaixo, escolha um município e veja como ele distribuiu seus gastos. Em cada gráfico, entenda quanto o que foi usado naquela área representa em relação ao orçamento executado total (em porcentagem) e qual é a despesa por habitante em seu município. Você também pode comparar os resultados de diferentes cidades que escolher.

**Sobre que cidade você quer saber?**

BUSQUE UMA CIDADE

COMPARAR COM OUTRAS CIDADES

PRONTO

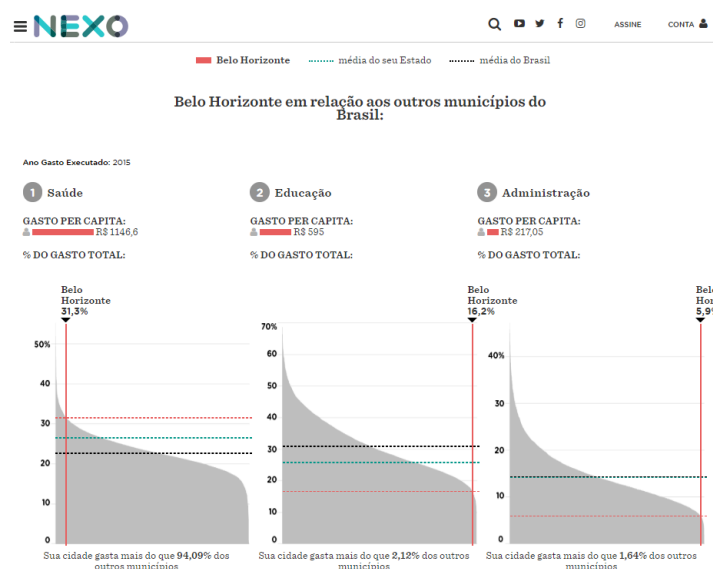
If only one city is selected, 11 graphics about its expenses are unveiled. The first is a colored bar chart that shows the spending by field of application (health, education, culture, education, agriculture, etc.).

Figure 26 - A bar chart shows the city's expenses by field of application



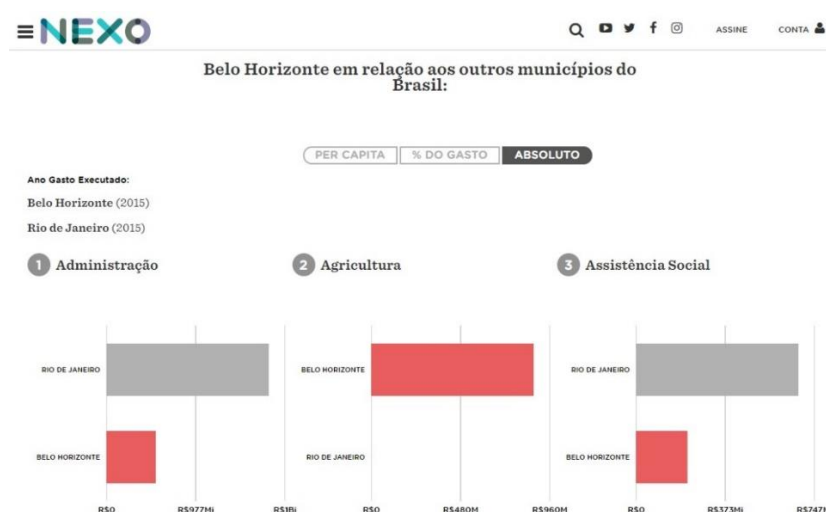
The other 10 are compound graphics, each one formed by area and line charts and representing data related to a specific field of application. These compound charts present the per capita expense of the city in Brazilian currency (R\$), the percentage of the spent in a field in relation to the whole executed city budget, and the state and country average expenses in the field. These are more complex graphics, that require a higher information visualization literacy. As Rodolfo Almeida mentioned, *Nexo's* audience is segmented and more specialized, what justifies such complexity level.

Figure 27 - Three examples of the compound graphics that show data of each field of budget application



When a second city is selected for comparison, 10 simple bar charts are showed, one for each field of budget application. There are three options for the reader to see the 10 graphics: per capita, spending percentage and absolute expenses. Then, the task *reconfigure* is present in this story.

*Figure 28 - Comparison between cities presents reconfiguration options*



In summary, the visualizations are the *center of the story*, mostly *interactive and exploratory*. The story is a narrative visualization presented in a *martini-glass structure* (Segel and Heer, 2010), with an author-driven introductory text and a following consultational and exploratory part.

*Table 11 - Summary of "Como sua cidade gasta dinheiro? Descubra e compare com outras", by Nexo*

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	no	yes	no	no	yes	no

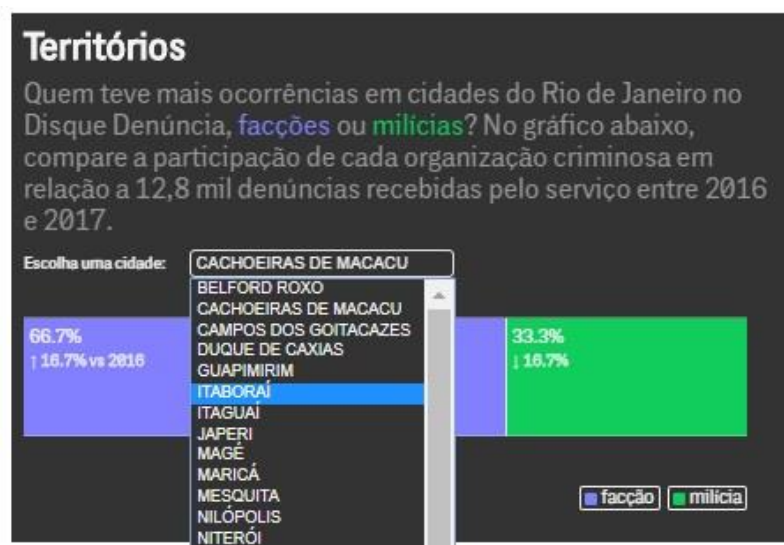
f) [Tá tudo dominado](#) - Volt Data Lab (for The Intercept Brasil), 5<sup>th</sup> Apr 2018

This story – produced by Volt Data Lab and published by The Intercept Brasil – is the fourth section out of six special reports about the death of Marielle Franco, a

councilor of Rio de Janeiro city, in March 2018. Marielle was known for denouncing the militia violence, and the suspicion arisen is that paramilitaries groups executed her. The report carried out by *Volt* team shows the growth of militias in Rio de Janeiro. The story mixes dozens of written paragraphs with seven graphics – three interactive and four static – built with data from a traditional denunciation hotline that receives anonymous information from Rio de Janeiro population. As there are no systematized databases about the growth of militias by the Rio de Janeiro state government, a detailed work of gathering, cleaning and analyzing the 12,800 anonymous calls related to drug dealers and paramilitaries activities received by the call center in 2016 and 2017 was needed.

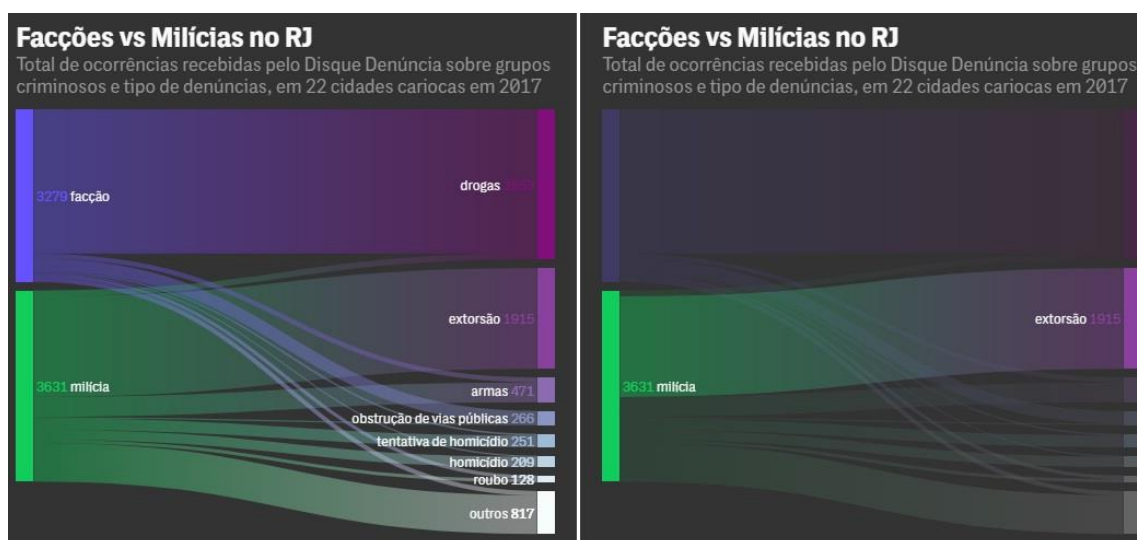
The first graph has an interactive option of *filtering* by city, from a drop-down menu. For each of the 22 municipalities comprised, a bar chart shows the percentage of denunciations related to criminal factions and to militia groups.

*Figure 29 - Interactive graphic with drop-down menu*



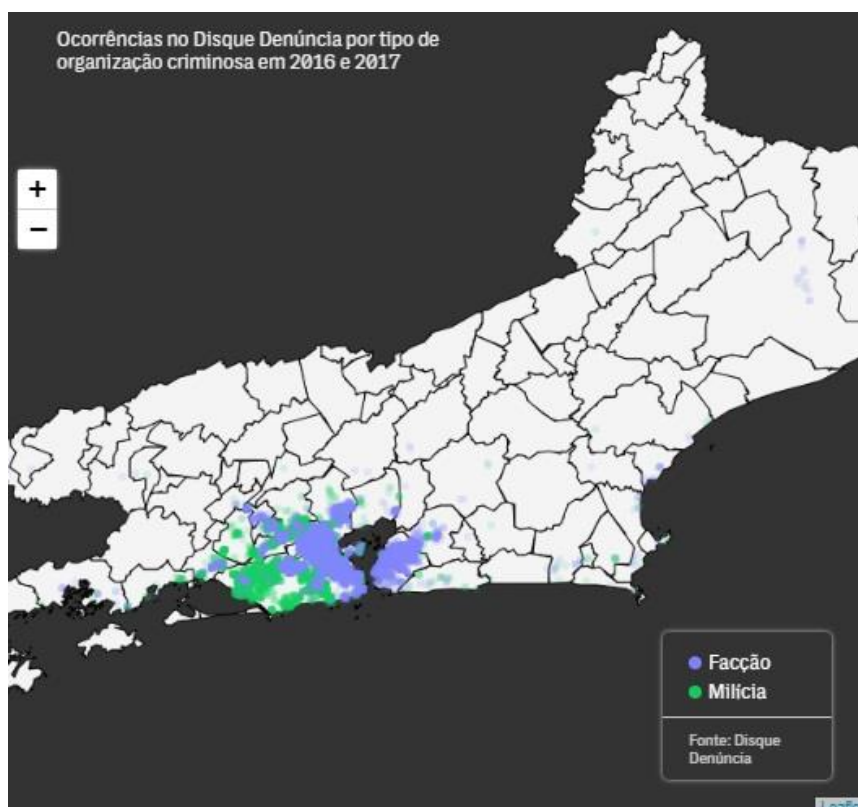
Another interactive graphic shows the total occurrences in 22 cities of Rio de Janeiro in 2017 received by the call center, divided between criminal factions and militia groups, on the left side, and by type of denunciations, on the right side. The interactivity takes place when the reader hovers the mouse on the graphic and it shows relationships among items. In other words, the interactive task presented in this chart is *relate*.

**Figure 30** – The interactive chart shows the relation between items when hovering the cursor on them.



An interactive map of the Rio de Janeiro state, in which is possible to *zoom* in and out, also geographically represents the distribution of criminal factions and militia groups. Even though there is no additional information on the zoomed map.

**Figure 31** - An interactive map shows the distribution of criminality by criminal factions and militia groups in the Rio de Janeiro State.



The other four static graphics are basic bar and line charts. In brief, all the visualizations are *part of the story*. As there is a balance between static and interactive graphics, the effort of resorting to interactivity is considered stronger. Thus, the visualizations are mostly *interactive* and *explanatory*.

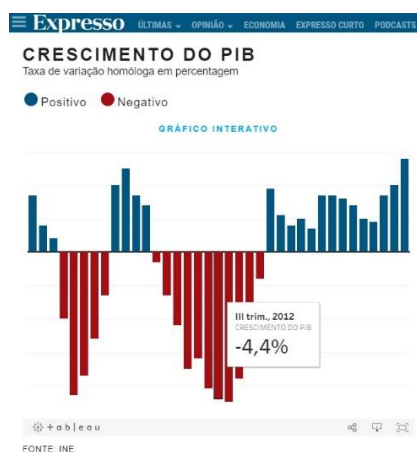
*Table 12 - Summary of "Tá tudo dominado", by Volt Data Lab (for The Intercept Brasil)*

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
4	-	yes	yes	no	yes	no	no

- g) [A nação enquanto não há remodelação: 12 gráficos para explicar o bom e o mau do país](#) - *Expresso*, 2<sup>nd</sup> Jul 2017

This *Expresso's* story presents 12 simple interactive graphics (bar, line and area charts) interleaved with blocks of written text. It presents an economic, political, judicial and demographic picture of Portugal. The main data sources are the National Statistical Institute, Bank of Portugal, Budget General Directorate, State General Account, Nature Conservation Institute and Forestry and Pordata (a database equipped with official and certified statistics about Portugal and Europe). The charts were built with [Tableau](#) and offer a basic interactive resource, that is hovering the cursor on elements to see more details, namely *details-on-demand*.

*Figure 32 - Graphics made with Tableau: the interactive task is detail-on-demand*



When clicking on elements, two bar charts present the options of “keep only” or “delete” (figure 33). Although it is not clear what this means or if it is intentional, it is a possibility of *reconfiguring* the graphic, showing only one year. However, if the reader wants to see both years again, it is necessary to refresh the page. That seems to be one of the limitations of using Tableau, as Sofia Miguel Rosa and Raquel Albuquerque had said during the interviews. Despite offering basic solutions of interactivity, the software is a viable alternative to circumvent the lack of a programmer.

Figure 33 - The chart can be reconfigured when clicking on "keep only" or "delete"



In short, the visualizations are *part of the story*, and despite their simplicity, they are *mostly interactive explanatory*.

Table 13 - Summary of "A nação enquanto não há remodelação: 12 gráficos para explicar o bom e o mau do país", by Expresso

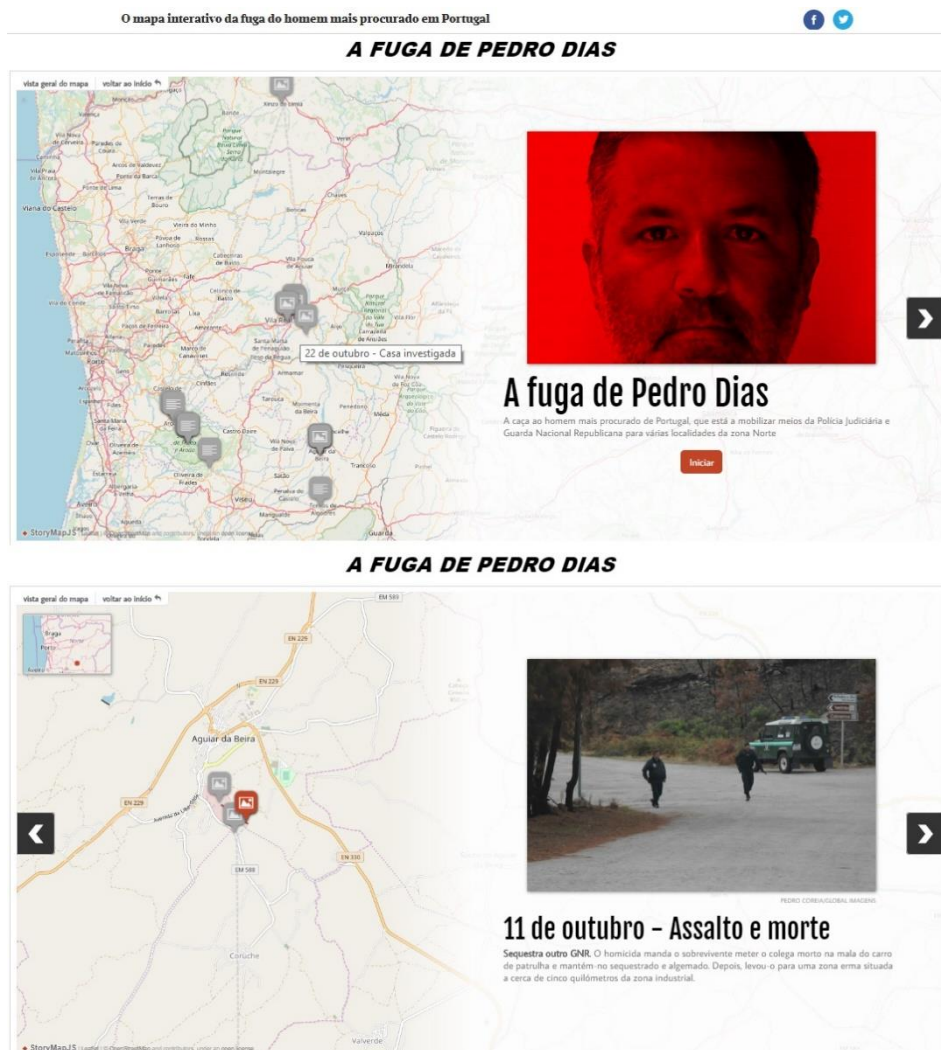
Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	no	no	yes	no	yes	no



- h) [O mapa interativo da fuga do homem mais procurado em Portugal](#) - *Jornal de Notícias*, 19<sup>th</sup> Oct 2016

This is an interactive slide show that uses a georeferenced map (StoryMapJS, as seen in the watermark), photos and small pieces of text to show the escape of Pedro Dias, a man hunted by the police. The reader has the option of clicking on the stepper-button (*narrate*), forward or backward, to see the chronological sequence of 19 slides, or browsing the map, clicking straight on elements plotted on it to get *details-on-demand*, not necessarily in a chronological order. Although there are no zoom buttons, it is possible to approximate the map using the mouse. On the right top corner of the slide show, there is a button to get an overview of the map and another to return to the beginning of the slide sequence.

**Figure 34** - Two interactive slides of the sequence that tells the story of Pedro Dias



It is clear that there is no data gathering, cleaning, and analysis behind the story, fact that does not preclude the use of interactive and visual resources to communicate a message. Hence, it is an example of a simple interactive solution used effectively to tell a story. In this case, the visualization is the *center of the story*, and it is *mostly interactive explanatory*. Therefore, it is a narrative visualization with an *interactive sequence structure*.

**Table 14** - Summary of "O mapa interativo da fuga do homem mais procurado em Portugal", by Jornal de Notícias

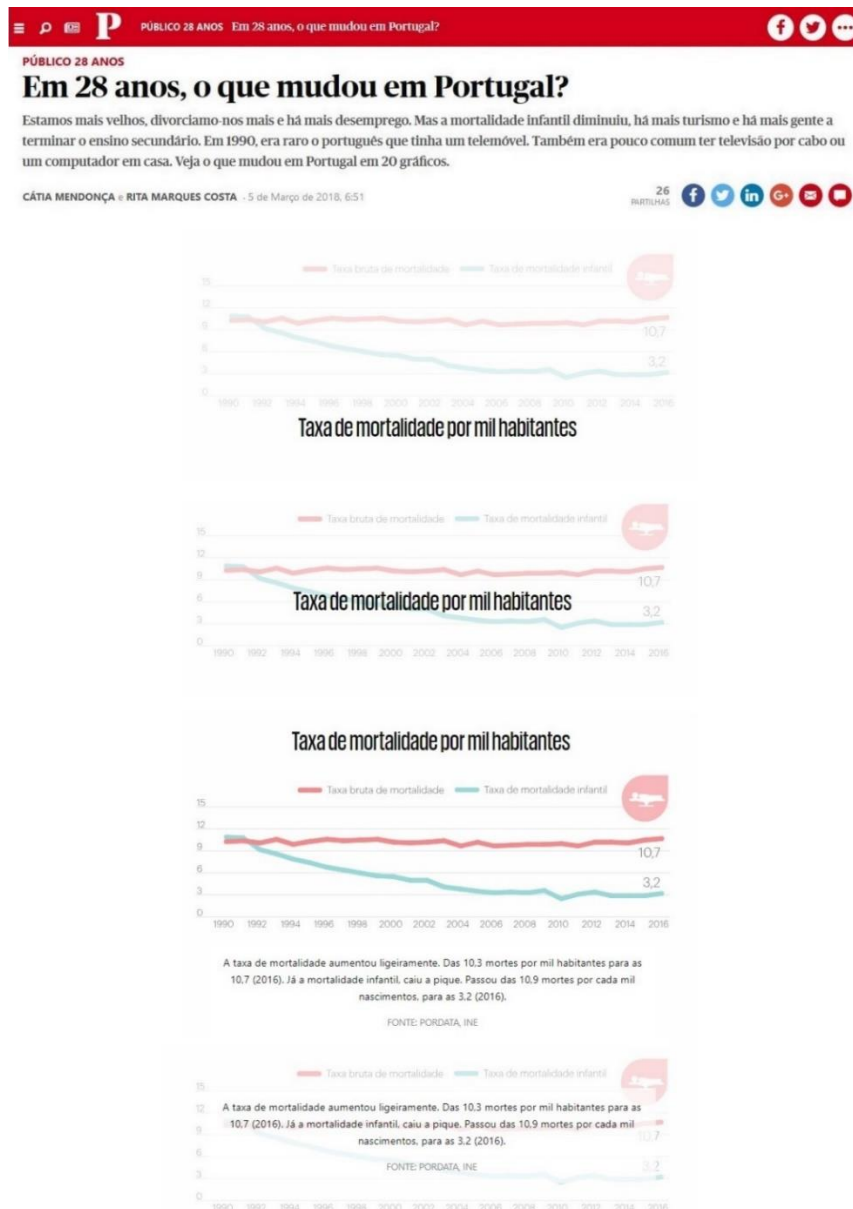
Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	yes	no	yes	no	no	yes

i) [Em 28 anos, o que mudou em Portugal?](#) - *Público*, 5<sup>th</sup> Mar 2018

This is a panorama of Portugal in the last 28 years, revealed through 20 graphics, small pieces of text and *scrollytelling*. The simple action of rolling down the screen reveals visualizations (bar and line charts) accompanied by explanatory messages about demography, economy, education, criminality, tourism, and consumption in the country. The datasets were drawn from the National Statistical Institute, Secretary-General of the Ministry of Internal Affairs, Justice Statistics Information System, Institute for Mobility and Transport, Budget General Directorate, Bank of Portugal, Institute of Cinema and Audiovisual, and Pordata (a database equipped with official and certified statistics about Portugal and Europe).

With the scrollytelling effect (*narrate* task), the charts appear in the background while the little blocks of text transverse the screen as a transparent layer on top of the visual information. The graphics fade in and out behind the text as the reader scroll down or up. It is a useful tool to transform traditional and static articles into a more attractive content other than the traditional article supplemented by static graphics.

Figure 35 - The scrollytelling creates an effect of interactive sequence, with the text running across the screen over the graphic background



Therefore, the visualizations are the *center of the story*, mostly *interactive* and *explanatory*. It is a narrative visualization presented in an *interactive sequence structure*.

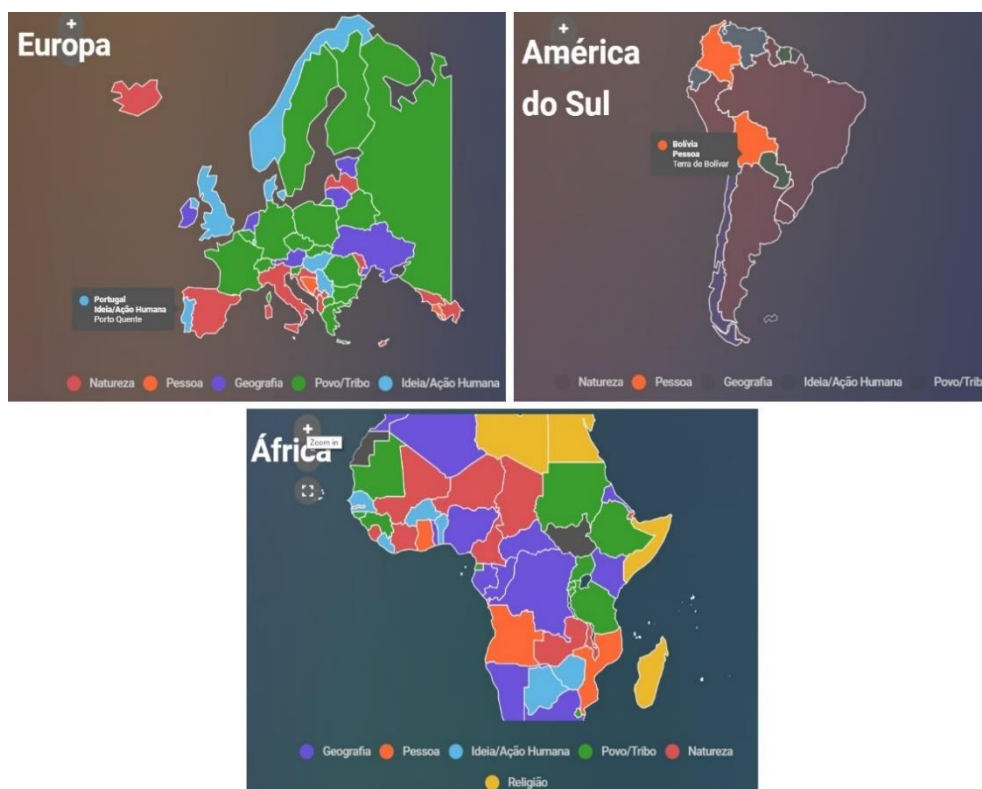
Table 15 - Summary of "Em 28 anos, o que mudou em Portugal?", by Público

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	no	no	no	no	no	yes

- j) [Mapa interativo. Sabe o significado literal do nome dos países?](#) - *Observador*, 21<sup>st</sup> Mar 2018

This story consists in a three-paragraph explanatory text followed by a colorful interactive map that shows the meanings of countries' names. Continents are represented separately, and countries are colored according to their names' etymological origin: religion, people or tribe, person, human idea or action, nature or geography. The database was structured by an Australian startup specialized in comparing credit cards, called Credit Card Compare, by crossed data from CIA, Oxford Dictionaries, and other etymological sources. The company published stylized static maps and the database in its [blog](#)<sup>51</sup>. Hence, great part of the data processing was ready. It seems that *Observador* translated the information, divided the name's origins into the six categories abovementioned and built the visualization with Infogram, an online data visualization platform that requires no coding skills. As the head of technology and design Leo Xavier highlighted during the interview, journalists are encouraged to use this tool, so that they can include interactive graphics in their articles by themselves.

Figure 36 - The three interactive tasks: details-on-demand, relate and zoom



<sup>51</sup> <https://www.creditcardcompare.com.au/blog/world-map-literal-translation-country-names/>

The interactive tasks present in this story are *details-on-demand*, by hovering the cursor on each country; *relate*, by clicking on one of the six origins (religion, people or tribe, person, human idea or action, nature or geography); and *zoom*, by clicking on the buttons (+ and -) available for each continent.

In this case, the visualization is the *center of the story*, and it is *mostly interactive explanatory*. It is a narrative visualization, presented in a *martini-glass structure*, since it begins with an author-driven explanation and opens to a reader-driven interactive part.

**Table 16** - Summary of "Mapa interativo. Sabe o significado literal do nome dos países?", by Observador

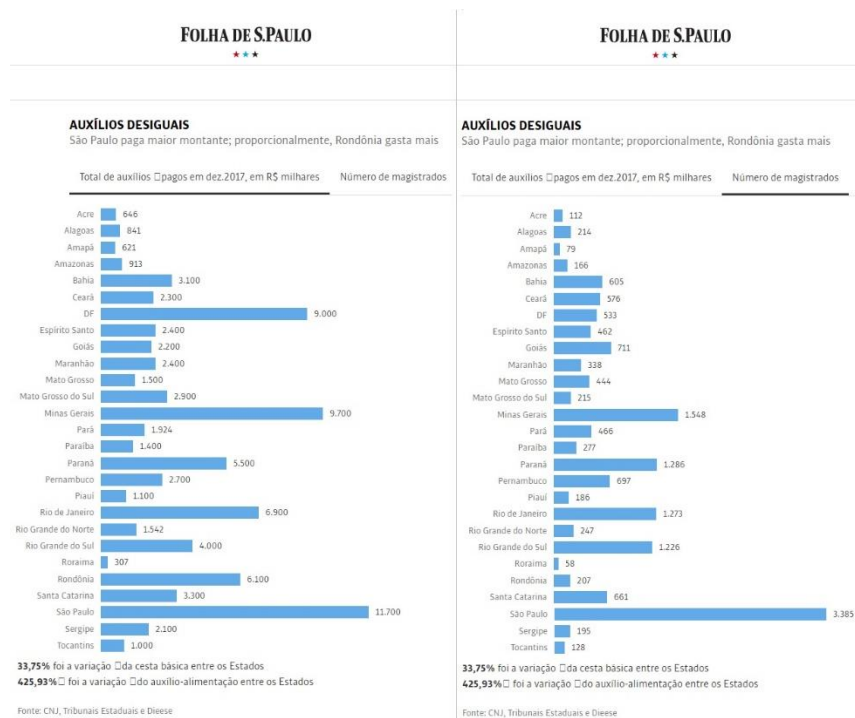
Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	yes	no	yes	yes	no	no

#### ***Brazilian stories related to the judiciary salaries***

- a) [Benefícios a juízes nos Estados têm disparidades e distorções](#) - Folha de S. Paulo, 18<sup>th</sup> Feb 2018

This is one of the data-driven stories produced in Brazil using the databases of the National Justice Council (CNJ) and regional courts with information about the judiciary salaries in each Brazilian member state. It is formatted as a classic article in what tables and graphics are illustrations for the written text. There are two long tables that hamper the reading, as they do not fit on the screen. Both present an effect of highlighting rows when hovering the cursor on them, which cannot be considered interactivity, as no new information is shown. There are also two bar charts: one is static and the other presents the option of *reconfiguring*, when clicking on a tab. However, this option is not clearly visible to the reader, as the tabs do not change colors. The only indication is a black underline, as seen below in figure 37.

Figure 37 - Each tab shows a reconfigured bar chart



Apart from this little interactivity, overall there is no great concern about a visual treatment of the data. Therefore, the visualizations are *part of the story*, and they are *mostly static and multiple*.

Table 17 - Summary of "Benefícios a juízes nos Estados têm disparidades e distorções", by Folha

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
3	-	no	no	no	no	yes	no

- b) [Com extras, 71% dos juízes do país recebem acima do teto de R\\$ 33 mil](#) - O Globo, 17<sup>th</sup> Dec 2017

In this story, the visualization is more complex and shows a concern with presenting an attractive content. Although the amount of text is similar to the area occupied by graphics, which are static, the visualization is only part of the story. It is also

noteworthy that there was an effort in merging different types of graphics: a map with a bubble chart (with size representing quantity), and a composite bar chart (where each bar displays multiple data points stacked in a single row or column). In summary, the visualization is *part of the story*, it is *static* and *compound*.

Figure 38 - O Globo's story resorts to a compound static graphic

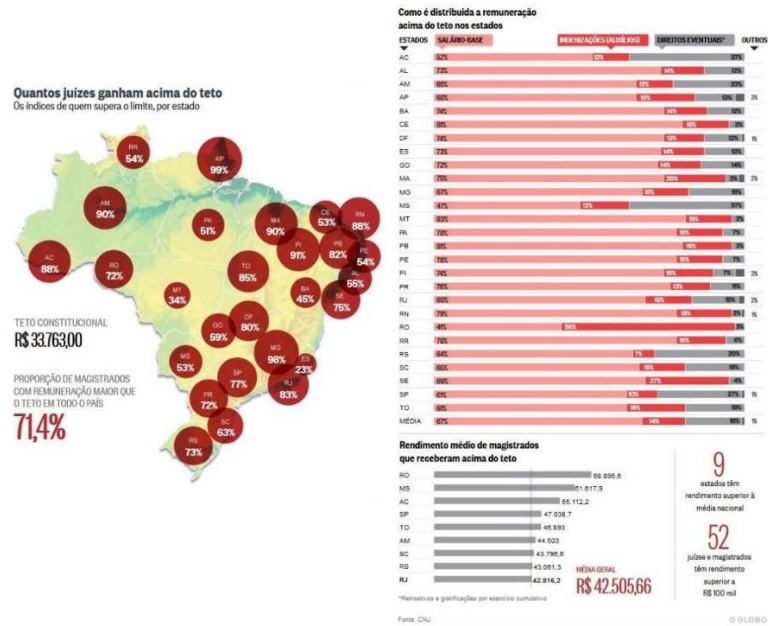


Table 18 - Summary of "Com extras, 71% dos juizes do país recebem acima do teto de R\$ 33 mil", by O Globo

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
3	-	no	no	no	no	no	no

c) [TJs gastam R\\$ 890 mi com 'penduricalhos' para juizes](#) - *Estadão*, 17<sup>th</sup> Dec 2017

This *Estadão's* story brings another example of a compound static information visualization. There are bubbles, with their size representing quantity, and a composite bar chart, with each bar displaying four different information points in a single row.



Figure 39 - Estadão's story presents a compound static infovis

## Ranking dos auxílios

Indenizações em forma de auxílios impulsionam o contracheque de um magistrado



\*O Amapá não foi contabilizado, pois entregou os contracheques depois dos outros tribunais  
Fonte: Conselho Nacional de Justiça (CNJ)

ESTADÃO

Summarizing, the visualizations is part of the story, it is static and compound.

Table 19 - Summary of "TJs gastam R\$ 890 mi com 'penduricalhos' para juízes", by Estadão

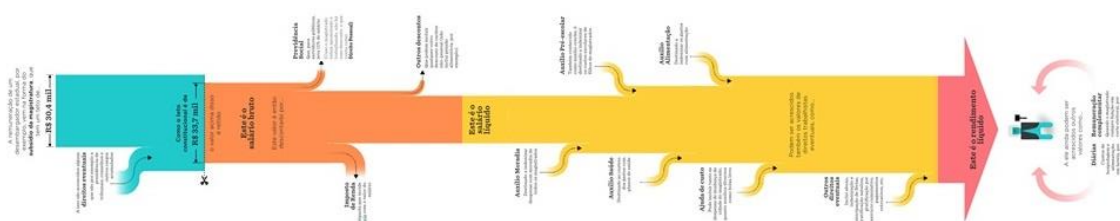
Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
1	-	no	no	no	no	no	no



d) [Quanto ganham desembargadores e juízes no Brasil](#) - Nexo, 21<sup>st</sup> Dec 2017

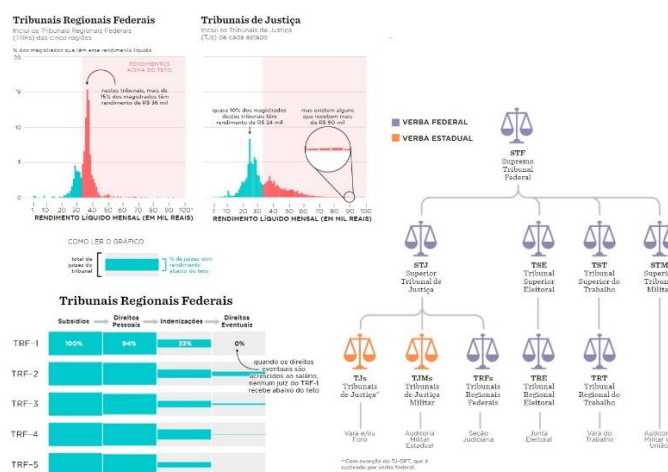
In this narrative visualization, as it is common in Nexo's data stories, graphics prevail over text. Written pieces are only annotations to guide the author. Despite being static, this visualization takes advantage of the scrolling down (in this case, it is not interactive, since the act of rolling the screen do not create any effect), presenting a long length Sankey diagram<sup>52</sup> that gives the notion of the amount of money the judicial magistrates earn, with all benefits and deductions. The complete overview of the story is available in figure 53, on [page xiv](#) of Appendix 1. For simplification, the illustration below was shortened and rotated.

**Figure 40** - A Sankey diagram representing judicial magistrates' wages (this illustration was shortened)



Apart from the Sankey diagram, there are more 15 graphics bar charts detailing information related to the several benefits and/or different justice courts, and one organizational chart representing a summarized structure of the Brazilian Judiciary.

**Figure 41** - Other kinds of charts used by Nexo



<sup>52</sup> Sankey diagram is a specific type of diagram to display flows and their quantities in proportion to one another.

As mentioned before, *Nexo* built more complex graphics because its audience is more specialized. The website also dispenses the traditional written articles, since its public has a higher information visualization literacy. However, when using an uncommon graphic, an explanation on how to read it is included.

Summarizing, the various visualizations are the *center of the story*, *static* and *multiple*. Hence, the story is a *static* narrative visualization.

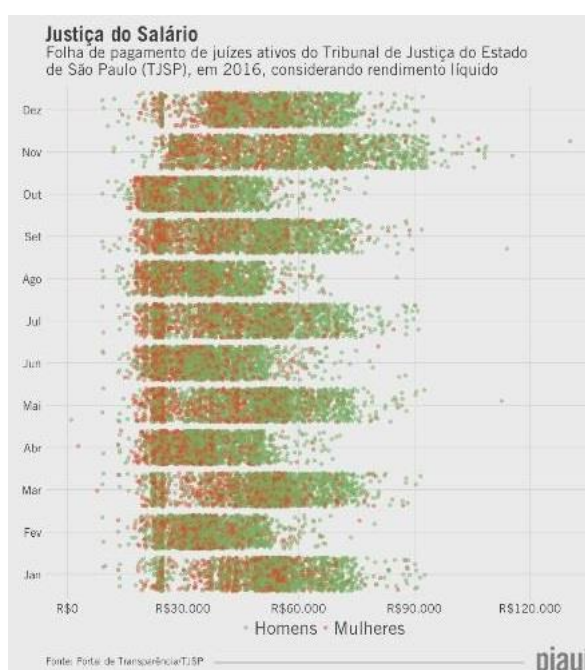
**Table 20** - Summary of "*Quanto ganham desembargadores e juizes no Brasil*", by *Nexo*

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
17	-	no	no	no	no	no	no

e) [Togas de ouro](#) - *Volt Data Lab* (for *Piauí Magazine*), 17<sup>th</sup> May 2017

The information visualization in this story has an evident central role, as the its position denounces: it is placed before the three-paragraph block of text, which, in its turn, presents supplementary information.

**Figure 42** - *Volt data Lab's* static compound visualization about the judicial salaries in São Paulo state



The story overall is shorter than the others analyzed above, because it is related to only one Brazilian member state, namely the São Paulo court. It resorts to a combination of a bar chart with a scatter plot chart (used to show the relationship between two sets of data) to reveal not only information about salaries in each month of 2016, but also the wage gap between men and women. In summary, the visualization is the *center of the story*, *static* and *compound*.

**Table 21** - Summary of "Togas de ouro", by Volt Data Lab (for Piauí Magazine)

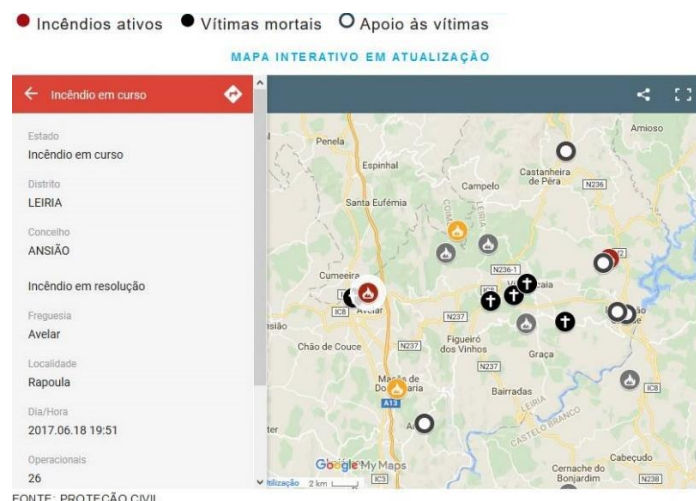
Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
1	-	no	no	no	no	no	no

### Portuguese stories related to the forest fires in 2017

- a) [A estrada da morte: onde o fogo matou – e onde ainda lava \(mapa interativo\)](#) - *Expresso*, 18<sup>th</sup> Jun 2017

The visualization, built using Google Maps, is itself the story. It shows points where the forest fires were still active, controlled or extinguished, as well as areas with fatalities and with support to the victims. There is no text beyond the title and subtitle. *Zoom in /out* and *details-on-demand* by clicking on elements are present.

Figure 43 - Example of the interactive task details-on-demand



Thus, the visualization is the *center of the story*, and it is *mostly interactive explanatory*. It is a narrative visualization with a *drill-down structure* (Segel and Heer, 2010) since it presents an overview first and then allows for additional details.

**Table 22** - Summary of "A estrada da morte: onde o fogo matou – e onde ainda lavra (mapa interativo)", by Expresso

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	Yes	no	yes	no	no	no

- b) [As freguesias nas listas de prioridade de limpeza de terrenos](#) - *Jornal de Notícias*, 16<sup>th</sup> Feb 2018

This is another example of a visualization that is itself the story. It is a map of Portugal divided into counties, showing all the first and second priority parishes for land cleaning, as a governmental strategy of fire prevention. Written text is only present in the title, subtitle and keys. The act of hovering on the counties unveils the number of parishes that will have lands cleaned. When clicking on them, parishes' names and priorities appear in a side box (*details-on-demand*). It is also possible to *zoom* in and out.

**Figure 44** - Details-on-demand and zoom are the interactive tasks in this story



Hence, the visualization is the *center of the story*, and it is *mostly interactive explanatory*. As a narrative visualization, it is presented in a *drill-down structure* (Segel and Heer, 2010), showing an overview first and then additional information.

**Table 23** - Summary of "As freguesias nas listas de prioridade de limpeza de terrenos", by Jornal de Notícias

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	yes	no	yes	no	no	no

- c) [Incêndio de Pedrógão Grande é o maior de sempre em Portugal](#) - *Público*, 19<sup>th</sup> Jun 2017

*Público* usually presents a more consistent production of data stories, but the use of interactivity is not the rule. This might happen because currently there is only a journalist in charge of data journalism. The data story about the biggest forest fire of the country is composed by a traditionally written article and four static visualizations with a supplementary role.

**Figure 45** - This *Público*'s story resorts only to static graphics and map

Os cinco maiores incêndios em Portugal

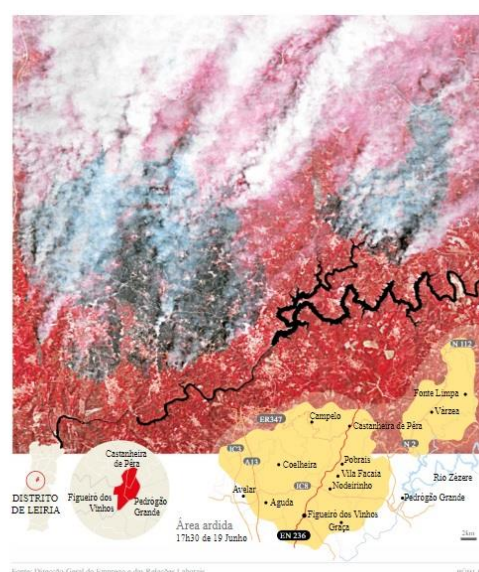
Freguesia	Concelho	Área ardida (ha)	Início	Extinção
Pedrógão Grande	Pedrógão Grande	25.969	17-06-2017	
Cachopo	Tavira	24.843	18-07-2012	20-07-2012
Ulme	Chamusca	22.190	02-08-2003	07-08-2003
Janarde	Arouca	21.910	08-08-2016	13-08-2016
Belver	Gavião	20.088	01-08-2003	06-08-2003



\* 1 de Janeiro a 15 de Outubro; \*\* 1 de Janeiro a 15 de Junho

Fonte: ICNF; EFFIS, Copernicus - Comissão Europeia; Deimos Imaging

PÚBLICO



Fonte: Direção-Geral do Emprego e das Relações Laborais

PÚBLICO

In this case, free visualization tools that do not require coding skills – such as Datawrapper, Tableau or Infogram – could have been used to include some interactivity. For example: details-on-demand would be useful to avoid the excess of information verified in the bar charts below. In addition, the yellow and red vertical bar charts could occupy a bigger area of the screen, with the option of reconfiguring the visualization, by clicking on a tab. Also, a free georeferenced map software would be a better option to show the burnt regions, with a zoom option.

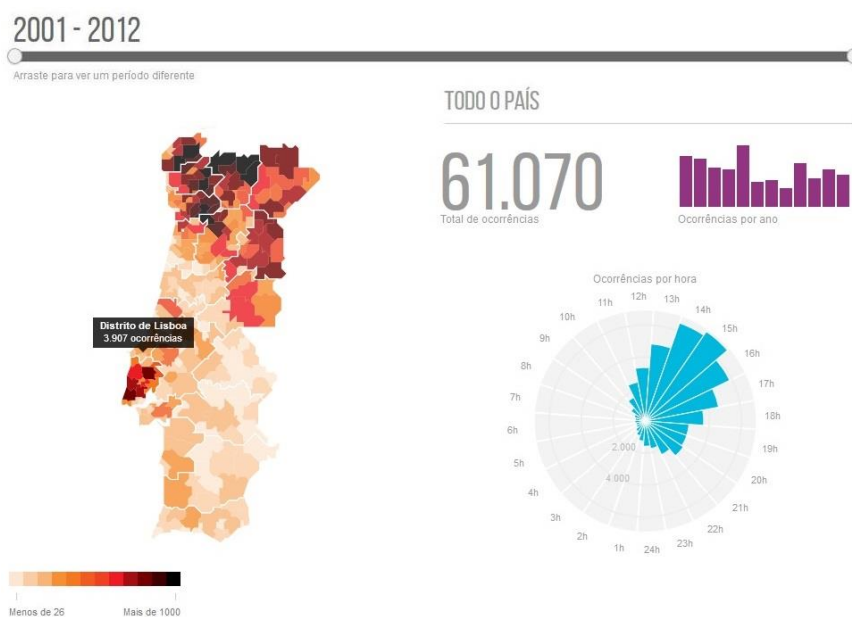
**Table 24** - Summary of "Incêndio de Pedrógão Grande é o maior de sempre em Portugal, by Público

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
4	-	no	no	no	no	no	no

For comparison purposes, the data story about forest fires [Doze anos de incêndios](#) produced by *Público* in 2012 as part of the REACTION research project (when there was a multidisciplinary team dedicated to data journalism) was also analyzed.

To build the interactive visualization – which comprise a Portugal map divided into counties, a vertical bar chart and a radial bar chart –, *Público* compiled in a single list the annual records of fires between 2001 and 2012, from the Institute of Nature Conservation and Forests. The story shows information about geographical location (by municipality), time and dates of the fires. There is no written text beyond the title, subtitles, and the explanation about how the database was built. The reader can *filter* the data by adjusting the year range in a draggable row. For each range, *reconfigured* visualizations are shown. All elements present *details-on-demand* by hovering the cursor on them. The map also offers the possibility of zooming in, when clicking on counties. In short, the visualizations are the *center of the story* and *mostly interactive exploratory*. It is a narrative visualization with a *drill-down structure* (Segel and Heer, 2010), presenting an overview first and then additional information when required.

**Figure 46** - Público's interactive story produced under the REACTION project



**Table 25** - Summary of "Doze anos de incêndios", by Público

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	no	yes	yes	no	yes	no

- d) [Veja o mapa interativo dos incêndios no centro do país](#) - *Observador*, 19<sup>th</sup> Jun 2017

This visualization was built using Leaflet, a JavaScript library for interactive maps. A short two-paragraph explanatory text precedes it. The map shows the location of the fires and offers the possibility of *zooming* and getting *details-on-demand*, as the phrase “hover the mouse on the map to get more information” indicates to the reader. After the map, there is also a simple infographic built with Infogram, with information on the number of firefighters, cars, and aircrafts used to fight the fires. In short, the visualization is the *center of the story*, and it is *mostly interactive explanatory*. It is a *drill-down structure* (Segel and Heer, 2010) narrative visualization, as it presents an overview followed by additional information.



**Figure 47** - The interactive tasks are zoom and details-on-demand



**Table 26** - Summary of "Veja o mapa interativo dos incêndios no centro do país", by Observador

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	-	yes	no	yes	no	no	no

e) [Área ardida em Portugal](#) - Frames (for Observador), 21<sup>st</sup> Apr 2018

This is an example of ‘frame’, a small static contextualizing chart included into *Observador* traditional writer articles. In this case, the bar chart shows data from the Nature Conservation Institute and Forestry about the burnt area (in hectares) between 2007 and 2017. The graphic is located into an article about a 16.2 million euros government investment to intervene in the hydrographic network of 57 counties affected by the fires in 2017. An arrow on the superior right corner gives the possibility to share the chart on social networks and a button for suggestions opens a message field. As all ‘frames’, the visualization *is part of the story*, and it is *static* and *simple*.



**Figure 48** - Example of a 'frame', its sharing options and box for suggestions.



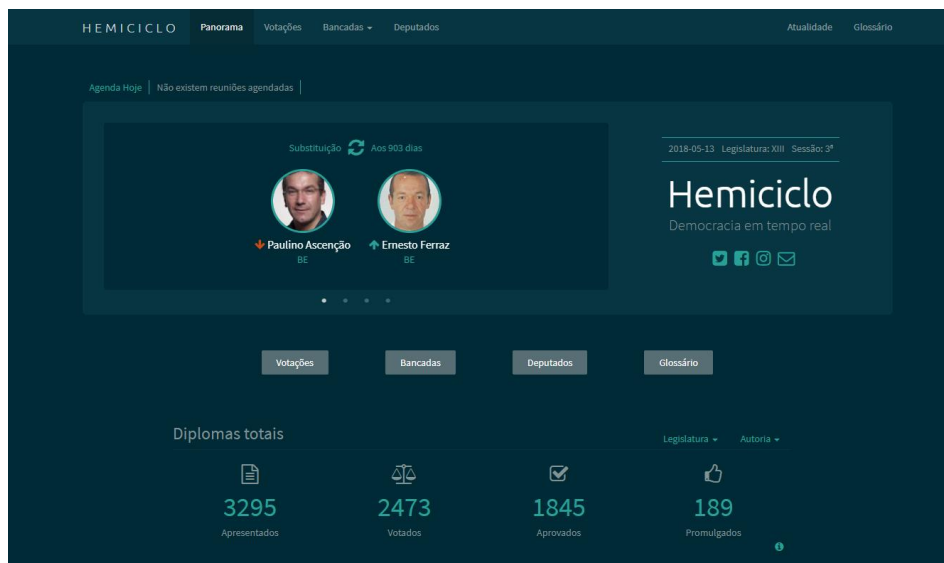
**Table 27** - Summary of "Área ardida em Portugal", by Frames (for Observador)

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
1	-	no	no	no	no	no	no

### **Hemiciclo evaluation**

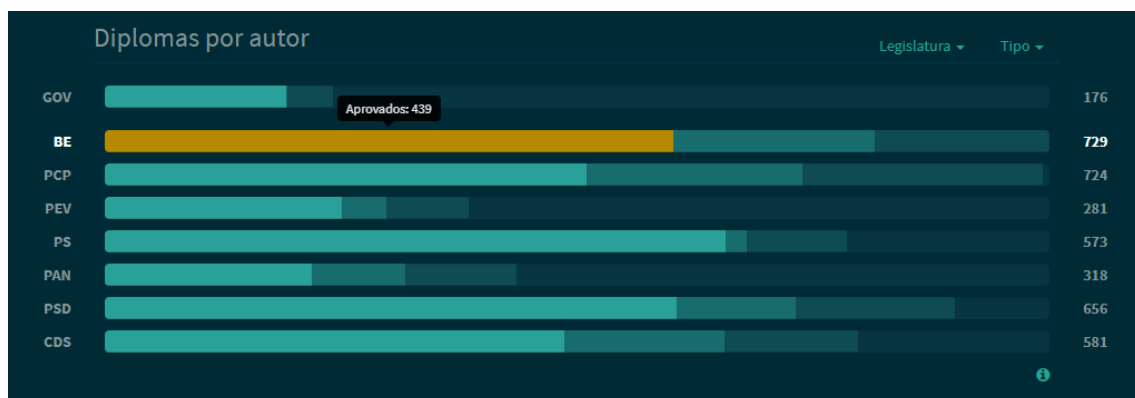
*Hemiciclo* is not a journalistic website, but it resorts to interactive visualizations to convey information about the Portuguese National Assembly, such as voting, political parties, deputies, bills, etc. Overall, written text is instrumentally used in labels, titles, subtitles, and annotations to guide the reader. The exception are sections such as actualities and glossary. The landing page presents a panorama (and this is its name on the superior menu) of the Parliament. Under a slide-show with links to articles (actualities), there are buttons that redirect the reader to other sections. Below, small clickable icons show numbers about presented, voted, passed, and enacted bills.

Figure 49 - Hemiciclo's landing page



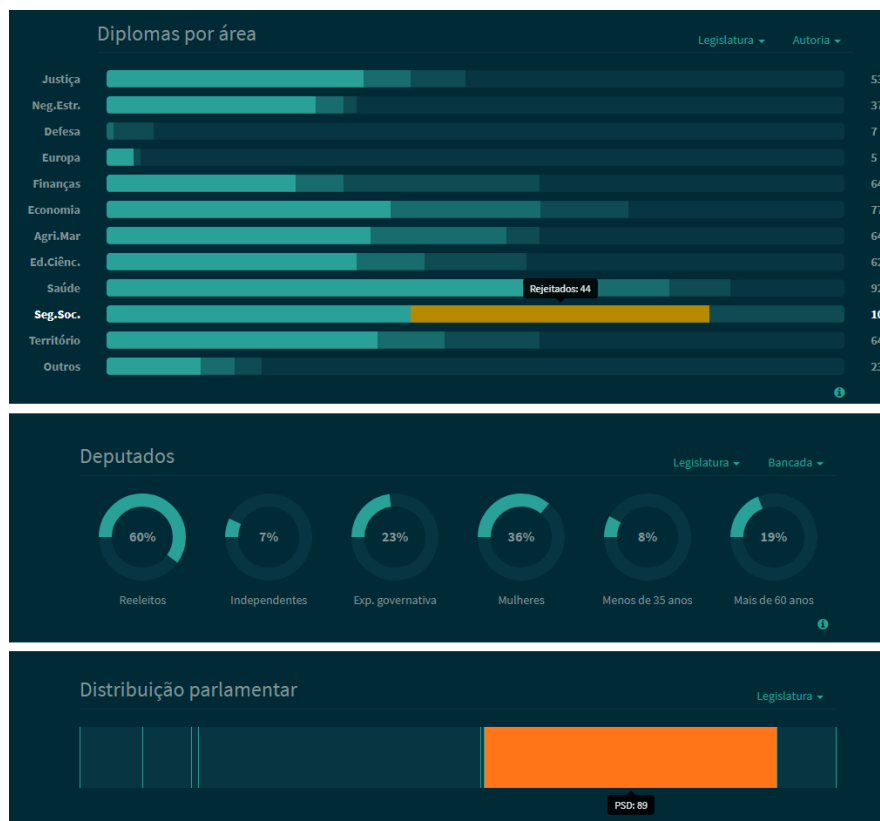
There is also an animated and interactive stacked bar chart that shows data about the bills by author (government and the several parties). It is possible to *filter* the information, and consequently *reconfigure* the chart, choosing the legislature and the type of the bill in two drop-down menus. By hovering the cursor (*details-on-demand*) on the bars, the reader gets the number of passed, rejected and not voted bills. When clicking on them, another section – voting – is unveiled. This is just one way to access this section. It can also be done via the superior menu, as well as by a button found in a privileged position in the first screen. The same happens with the sections ‘parties’, ‘deputies’ and ‘glossary’ (figure 49).

Figure 50 - Interactive stacked bar chart, with drop-down menus



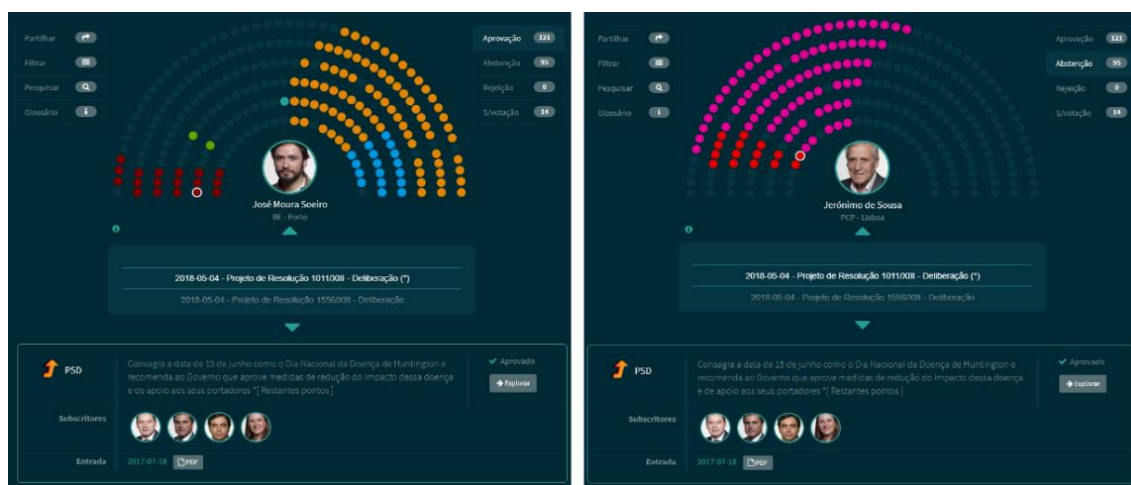
Further down, another stacked bar chart presents the bills divided by area, such as health, finance, defense, economy, justice and so forth. The same interactive features are present. Below, a set of six donut charts show percentage data about the deputies, such as the reelected, the independent, age and gender. On the bottom of the page, an interactive (supports hovering and clicking) stacked bar shows the political benches' distribution in the National Assembly.

**Figure 51** - Other interactive charts on the Hemiciclo's landing page



In the section voting, the visualization is a hemicycle (and that is the website's name, in Portuguese) representing the benches configuration of the National Assembly. The colors vary according to the political bench. Each dot symbolizes a deputy. It is possible to choose a bill (*filter* and *reconfigure*) from a list and see who voted for and against it, the abstentions and absences.

**Figure 52** - Visualization of a voting: in the left, the colored dots show who voted for and in the right, the colored dots show the abstentions



Deputies' photo, party, and county of origin are shown when hovering the mouse on the dot. Clicking on it redirects to a page with a biography and bills that the deputy has voted. Another way to access deputies' information is by clicking on the superior menu and filling up the search field with the deputy's name or filtering by legislature, bench and county of origin. The hemicycle visualization, the stacked bar chart, and pictorial information are also present in the section about the political benches, with the same interactivity resources. In brief, the website *Hemiciclo*, although is not a journalistic piece, is a good example of communicating messages behind complex datasets. The visualizations have a *central role* and are *mostly interactive exploratory*.

**Table 28** - Summary of Hemiciclo

Static Infovis	Dynamic Infovis						
	Noninteractive (animation)	Interactive					
		Zoom	Filter	Details-on-demand	Relate	Reconfigure	Narrate
-	multiple	yes	yes	yes	no	yes	no

Table 29 - Summary of the qualitative analysis

DATA STORY	INFOVIS		MOSTLY STATIC			MOSTLY DYNAMIC		
	PART OF THE STORY	CENTER OF THE STORY	SIMPLE	MULTIPLE	COMPOUND	MOSTLY NONINTERACTIVE	MOSTLY INTERACTIVE	
							EXPLANATORY	EXPLORATORY
Música muito popular brasileira - Folha de S. Paulo								
Barcelona, capital catalã, teve maior proporção de votos contrários à separação - O Globo								
Em queda, roubos de carros migram para a periferia - Estadão								MARTINI-GLASS
As mortes violentas mês a mês no país - G1								DRILL-DOWN
Como sua cidade gasta dinheiro? Descubra e compare com outras - Nexo								MARTINI-GLASS
Tá tudo dominado - Volt Data Lab (for The Intercept Brasil)*							*	
A nação enquanto não há remodelação: 12 gráficos para explicar o bom e o mau do país – Expresso*							*	
O mapa interativo da fuga do homem mais procurado em Portugal - Jornal de Notícias							INTERACTIVE SEQUENCE	
Em 28 anos, o que mudou em Portugal? - Público							INTERACTIVE SEQUENCE	
Mapa interativo. Saiba o significado literal do nome dos países? - Observador							MARTINI-GLASS	
Benefícios a juizes nos Estados têm disparidades e distorções - Folha de S. Paulo								
Com extras, 71% dos juizes do país recebem acima do teto de R\$ 33 mil - O Globo								
TJs gastam R\$ 890 mil com 'penduricalhos' para juizes - Estadão								
Quanto ganham desembargadores e juizes no Brasil - Nexo								
Togas de ouro - Volt Data Lab (for Piauí Magazine)								
A estrada da morte: onde o fogo matou – e onde ainda lava (mapa interativo) - Expresso							DRILL-DOWN	
As freguesias nas listas de prioridade de limpeza de terrenos - Jornal de Notícias							DRILL-DOWN	
Incêndio de Pedrógão Grande é o maior de sempre em Portugal - Público								
Doze anos de incêndios - Público								DRILL-DOWN
Veja o mapa interativo dos incêndios no centro do país - Observador							DRILL-DOWN	
Área ardida em Portugal - Frames (for Observador)								
Homicídio								DRILL-DOWN

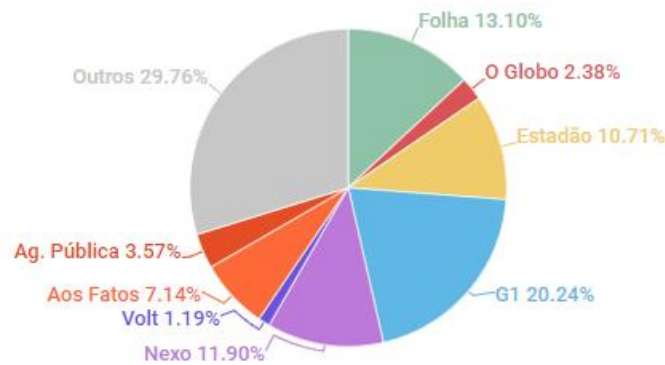
\* Despite being interactive, the visualizations in “Tá tudo dominado” and “A nação enquanto não há remodelação” are part of the story, which means they are not narrative visualizations. Thus, the Segel and Heer’s classification (martini-glass, interactive sequence and drill-down) is not applicable for those cases.

### 3.4 – Quantitative assessment of the DDJBR list

[DDJBR](#), created by *Volt Data Lab*, is a list of data journalism stories considered exemplary in Brazil, curated by journalists from several Brazilian media outlets. By the time of this dissertation writing, three editions of the list were available: March, April, and May. In Portugal, there is no similar list. Thus, the quantitative evaluation was conducted only with Brazilian stories. Therefore, the comparison between the two countries is strictly qualitative.

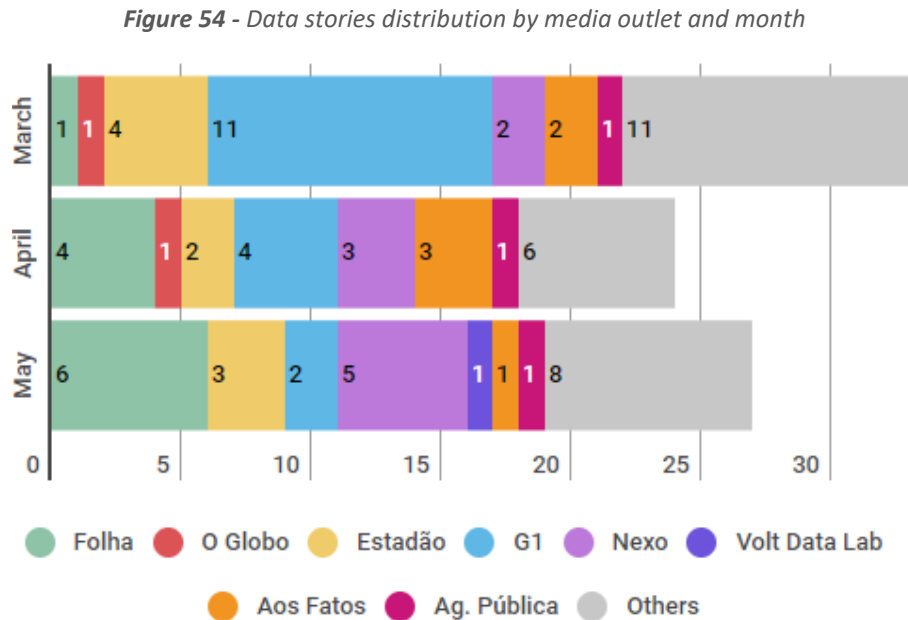
In total, 84 stories from 24 publications were assessed. The visualizations produced in video format were not considered, as the proposed taxonomy was developed taking interactivity into account. The pie chart below shows the media outlets qualitatively analyzed in the previous section and two more vehicles that published at least three data stories during the period between March and May. Those with one or two stories (except *O Globo* and *Volt Data Lab*, which appeared with two stories and one story respectively) were reunited into the category ‘others’.

Figure 53 - Data stories distribution by media outlet

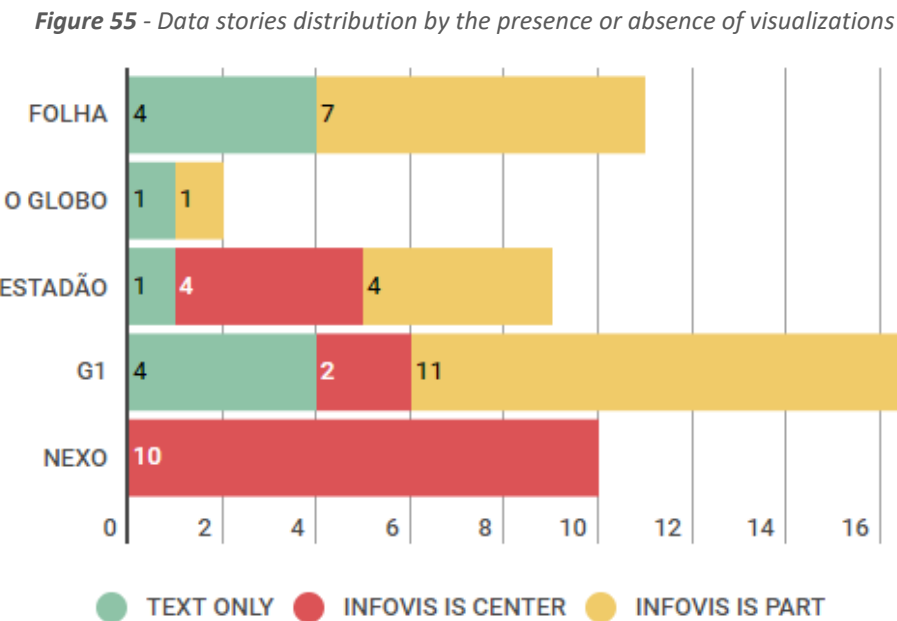


*G1* accounts for 17 data stories (20.24%), *Folha* appears with 11 (13.10%), *Nexo* has 10 (11.9%), *Estadão* responds for nine stories (10.71%), *Aos Fatos* is responsible for six (7.14%), *Agência Pública* has three (3.57%), *O Globo* published two (2.38%) and *Volt Data Lab* has one story on the list (1.19%). The little participation of *O Globo* can be justified by the priority given to fact-checking over data journalism. *Volt Data Lab* low presence seems to be due to its partnership with other media outlets and organizations, deep investigations and ongoing projects.

Considering these eight media outlets and the three editions of the DDJBR list (March, April, and May), the distribution of the stories is the following, as shown in the stacked bar chart below.



Considering the five Brazilian media outlets previously accessed, stories with no visualization, stories in what the visualization has a central role, and stories that have a supplementary visualization were divided as displayed below.

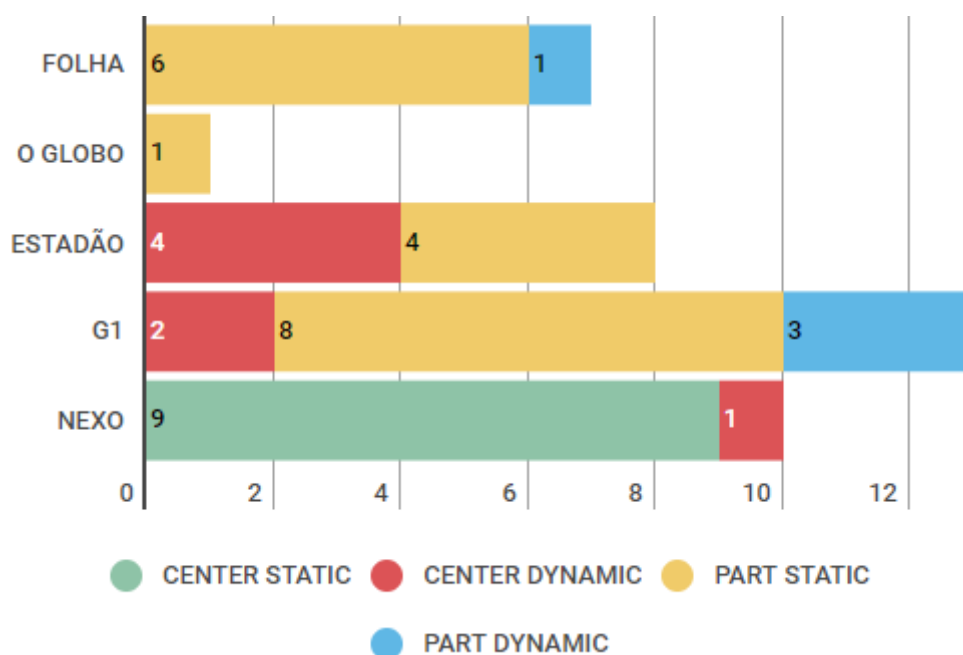


Among the 11 stories produced by *Folha de São Paulo*, there are no narrative visualizations (when the visualizations are the center of the story), four are composed by text only and seven present illustrative visualizations. In other words, all *Folha's* data stories are quite traditional when it comes to the format. Fábio Takahashi's responses are in line with those results. He declared that visualizations are not a goal and the data stories are produced to be apparently similar to the traditional articles.

*G1's* stories are mostly traditional as well. The massive production in number is because of the deep investigations projects – as Thiago Reis related – that often result in several stories covering the same subject. As an online-only vehicle, visualizations are used more often, but still very dependent on written articles. *Estadão* presents a more balanced production when it comes to visualizations being the center or part of the story. Visualizations are not a priority, but are used more often, even though sometimes with a printed newspaper mindset.

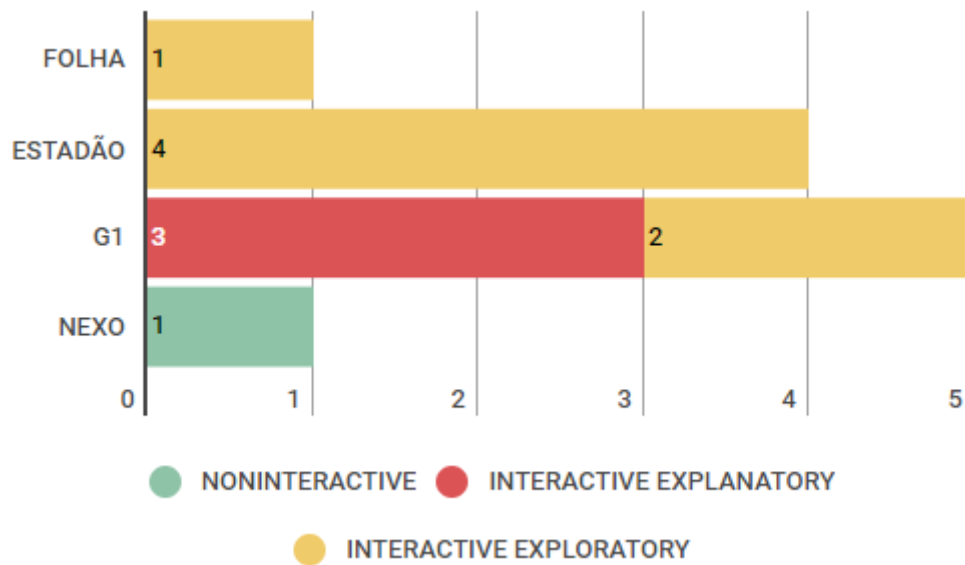
*Nexo* appears in prominence, with 100% of narrative visualizations, in accordance with Rodolfo Almeida described: the priority is the visual message, whereas and text has as a guideline role. However, interactivity could be more explored, as the following charts reveal. Only one story is dynamic, and it is noninteractive (animation).

Figure 56 - Stories with visualizations

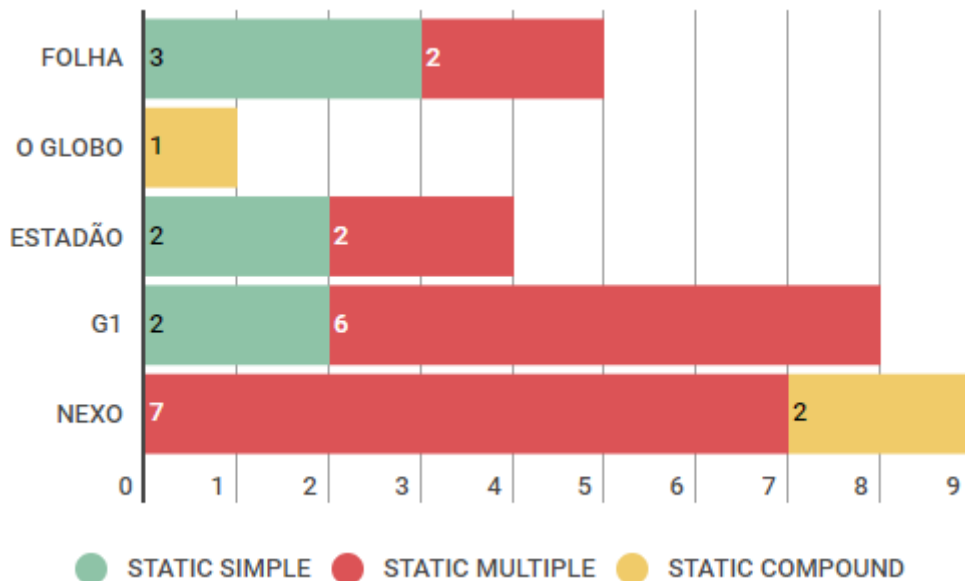




*Figure 57 - Stories with dynamic visualizations*



*Figure 58 - Stories with static visualizations*



### **3.5 – Qualitative comparative analysis**

The literature review of data journalism's research both in Brazil and Portugal, as well as the analysis carried out for this dissertation, revealed a recent interest in the field. However, this it seems to be happening in a cluttered and non-systematized way. It appears that Brazil's efforts to include data analysis and visualization in its publications

is ahead of Portugal's. However, the two countries still have a lot to evolve, mainly when it comes to the mainstream media. Likewise, the related academic production in both countries is sparse and embryonic, with Brazilian scholars showing more interest for the field than the Portuguese counterparts.

The fundamental factor that places Brazil forward to Portugal is the presence of multidisciplinary teams, dedicated only (or at least mainly) to data journalism. All Brazilian media outlets analyzed institutionally recognize the data journalism practice within their newsrooms. In Portugal, *Expresso* and *Público* are the vehicles that officially acknowledge this practice, but the companies have only one professional each in charge – not exclusively – of data journalism.

This influences other aspects that enlarge the gap between the two countries, such as workflow, technology, experiences and knowledge exchanges, training supply, and academic research. In Brazil, the process of gathering, cleaning, analyzing, and communicating data is more systematized, as the teams include programmers and/or designers. Such multidisciplinary not only entails the use of a more specialized technology, but also stimulates the creation of groups and events – such as the DDJBR WhatsApp group and the conference Coda.Br – to discuss themes within data journalism scope and the emergence of training courses – like the Abraji and Escola of Dados workshops and boot camps – to bridge the skills gap. Finally, all those elements foster the development of the related research.

Apart from *Nexo*, *Estadão*, and somehow *G1*, visualization is not in the focus of Brazilian data journalism professionals. They seem to value more the data processing rather than its visual communication. Apparently, the opposite is happening in Portugal. The media outlets have been including narrative visualizations in their publications, even when there is no data processing behind it or when the vehicle members do not perceive that practice as data journalism. Therefore, the data journalism workflow in both countries often ends up incomplete. In summary, the issues that Brazil's publications have are mainly in the visualization stage, whereas Portugal still has a long way to go towards more consistent and complete incorporation of data journalism in the newsrooms. The key seems to be investing in structuring dedicated multitasking teams.

## CONCLUSION

The amount of available data has astonishingly grown in the last years. This fact combined with the expanding storage capacity, open source accessible software capable of managing data that once required supercomputers, and transparency culture raised the need to process, structure and make sense of such informational ocean. Diverse areas are taking advantage of the data analysis. Journalism has been including other fields – statistics, design, and programming – to the traditional newsmaking, leading to the emergence of a new modality called *data journalism*. In the last decade, The Guardian and The New York Times, among few others, consolidated the novel sub-field that can be understood as the practice of extracting useful information from datasets and communicating it through visualizations and interactivity.

Data journalism is also described as a process that encompasses at least four stages – gathering, filtering, analyzing and communicating – with the latter being developed preferably in digital media in order to include interactive visualizations. Otherwise, data is used solely to find or enhance stories, which is not a novelty in journalism. Nevertheless, this is the instrumental use of datasets that this study has verified in some media outlets both in Brazil and Portugal, even in vehicles that already have multidisciplinary teams dedicated only to data journalism. *Folha de São Paulo*, for example, recently incorporated a team of four programmers that work mainly in data process and analysis, a practice that is closer to computer-assisted reporting (CAR) than to data journalism. The same was verified in some stories by *G1*, despite its efforts to occasionally include interactive visualizations.

This use of data cannot be considered a worse approach. As Lorenz (2010) suggests, CAR is still very relevant for investigative journalism. However, it is more a technique to be used within the data journalism practice rather than a process that affects the traditional journalism workflow in a fundamental way. Evidently, including interactivity and visualization is not mandatory for good data stories, particularly because such resources must be used only when they add value to the communication of the message. Nevertheless, there are several reasons to considerate the use of visualizations in data journalism. Citing a few, they are attractive and attention-grabbing, convey information in a simpler and more compelling way while amplifying

cognition (as the human brain has facility to process visual information), and provide transparency about the reporting process when they are interactive explorative.

*Nexo* seems to make an appropriate use of visualizations. The website appears, thus, as a leading outlet in Brazil when it comes to data journalism, even though the use of interactivity is limited. It publishes at least three data stories per week, which are essentially narrative visualizations (information visualizations that are themselves the storyteller). A plausible reason for this preponderance is the fact of *Nexo* being an online-only publication created parallelly to the mainstream media. Such scenario seems to favor a tech-savvy mindset, decoupled from the traditional media *modus operandi*. Similarly, the Portuguese vehicle *Observador* – also online-only and apart from the mainstream media groups – presents an innovative automated model to include minor charts to its articles, an approach that seems to accomplish its function of adding relevant graphics related to the articles' subjects. In addition, both websites are quite recent enterprises that have in common the fact that they chose to hire multitasking professionals from the beginning.

*Estadão*, even going through a shortage of workforce, is still a reference in Brazil, given its pioneering initiative of including programmers into a data-journalism team. Its benchmark appears in the flagship story analyzed for this dissertation and in the spreading of the data journalism culture across its newsroom.

In Portugal, *Público's* path reveals how important is having a team fully dedicated to data journalism. The difference of the current production, with just one data journalist (but not dedicated only to data work), and the practice verified in 2012-2013 during the REACTION research project (with an exclusive multidisciplinary staff) is remarkable. Interestingly, all of the Portuguese outlets assessed still need to realize the importance of structuring cross-disciplinary teams if they want to master the data journalism practice.

Although *Expresso* has a designated data journalist, its practice is very limited because of resource scarcity. Beyond the few data stories, the vehicle has a section in its website called 2:59, which consists in nearly-three-minute videos that often includes data work. They were not analyzed because the video format does not comprise interactivity and, thus, it is beyond the scope of this dissertation.

However, the combination of video and data journalism seems to be an interesting theme for future research, as it has also appeared in the DDJBR list and are often used by some famous outlets, such as *VOX*, an American news and opinion website owned by Vox Media. Another valuable approach for future research is the data journalism reception perspective, focusing on how readers perceive and interact with narrative visualizations. For that matter, it is worth including the data journalism perspective into the current visual literacy studies, as the media plays a vital educational role in making audiences familiar with information visualizations.

A possible limitation of the present study is perhaps not having directly compared the Brazilian and Portuguese media outlets with a flagship publication, such as *The Guardian* and *The New York Times*. However, the access to foreign flagship vehicles' staff usually is not that straightforward and would require more time and resources to accomplish. Moreover, literature review compensates such gap, as the most prominent media outlets have been exhaustively studied. Possibly, it also would have been useful to have carried a temporal analysis of the chosen media in Brazil and Portugal. Nevertheless, this would require a thorough investigative work, as the websites (except *Nexo*) do not have a consistent periodicity to publish data journalism stories, neither separated sections for such content. *Folha*, for example, explicitly prefers not to separate data stories from the others, as declared Fábio Takahashi: "the idea, so far, is not to distinguish the data journalism material from what is investigative journalism or any other matter. So, we do not use a 'data journalism stamp'". Therefore, to find data stories by period would be a tough task, not to mention the risk of leaving something out. Hopefully, this scenario will be improved soon so that more detailed investigations can be carried out in a not-too-distant future.

## REFERENCES

- Alexandre, I.A.R., 2014. *Jornalismo de Dados: o estado da arte nos jornais generalistas diários em Portugal*. Nova University of Lisbon, Lisbon.
- Barlow, M., 2014. *Data Visualization: A New Language for Storytelling*. O'Reilly Media, Sebastopol, CA.
- Bastos, H., 2010. *Ciberjornalistas em Portugal: práticas, papéis e ética*, Media e jornalismo. Livros Horizonte, Lisboa.
- Berret, C., Phillips, C., 2016. *Teaching Data and Computational Journalism*. Rosemont Press, New York.
- Blank, S., 2013. Why the Lean Start-Up Changes Everything [WWW Document]. Harv. Bus. Rev. URL (accessed 4.23.18).
- Börner, K., Maltese, A., Balliet, R.N., Heimlich, J., 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Inf. Vis.* 15, 198–213.  
<https://doi.org/10.1177/1473871615594652>
- Bounegru, L., 2012. Data Journalism in Perspective, in: Bounegru, L., Gray, J., Chambers, L. (Eds.), *The Data Journalism Handbook*. O'Reilly Media, Sebastopol, CA.
- Boy, J., Detienne, F., Fekete, J.-D., 2015. Storytelling in Information Visualizations: Does it Engage Users to Explore Data? *ACM Press*, pp. 1449–1458.  
<https://doi.org/10.1145/2702123.2702452>
- Bradshaw, P., 2012. What Is Data Journalism?, in: Bounegru, L., Gray, J., Chambers, L. (Eds.), *The Data Journalism Handbook*. O'Reilly Media, Sebastopol, CA.
- Bradshaw, P., 2010. How to be a data journalist [WWW Document]. *The Guardian*. URL <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide> (accessed 9.29.17).
- Brehmer, M., Munzner, T., 2013. A Multi-Level Typology of Abstract Visualization Tasks. *IEEE Trans. Vis. Comput. Graph.* 19, 2376–2385.  
<https://doi.org/10.1109/TVCG.2013.124>
- Cairo, A., 2017a. Nerd journalism: How data and digital technology transformed news graphics. Universitat Oberta de Catalunya. Internet Interdisciplinary Institute (IN3), Barcelona.
- Cairo, A., 2017b. Uncertainty and graphicacy: How should statisticians, journalists, and designers reveal uncertainty in graphics for public consumption?, in: *Power from Statistics*. Brussels.
- Card, S.K., Mackinlay, J.D., Shneiderman, B., 1999. *Readings in information visualization: using vision to think*, The Morgan Kaufmann series in interactive technologies. Morgan Kaufmann Publishers, San Francisco, Calif.
- Coddington, M., 2015. Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digit. Journal.* 3, 331–348.  
<https://doi.org/10.1080/21670811.2014.976400>
- Cohen, S., 2012. Using visualizations to Tell Stories, in: Bounegru, L., Gray, J., Chambers, L. (Eds.), *The Data Journalism Handbook*. O'Reilly Media, Sebastopol, CA.
- Cox, M., Ellsworth, D., 1997. Application-controlled demand paging for out-of-core visualization. Presented at the 8th conference on Visualization '97, IEEE

- Computer Society Press Los Alamitos, CA, USA ©1997, Phoenix, Arizona, USA, pp. 235–244.
- Craft, B., Cairns, P., 2005. Beyond Guidelines: What Can We Learn from the Visual Information Seeking Mantra? IEEE, pp. 110–118.  
<https://doi.org/10.1109/IV.2005.28>
- Cunha, M.P., 2017. Tecnologia, velocidade e jornalismo: Tensões e possibilidades da produção de reportagem em ambientes digitais. Presented at the 14º Fórum de Pesquisa Cáspes Líbero, São Paulo.
- Deuze, M., 2001. Online journalism: Modelling the first generation of news media on the World Wide Web. First Monday 6. <https://doi.org/10.5210/fm.v6i10.893>
- Diakopoulos, N., 2011. A Functional Roadmap for Innovation in Computational Journalism [WWW Document]. Nick Diakop. URL <http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/> (accessed 11.15.17).
- Faustino, P., 2004. The press in Portugal: changes and trends. MediaXXI/Formalpress, Lisboa.
- Figueiras, A., 2016. How to tell stories using visualization: strategies towards Narrative Visualization. Nova University of Lisbon, Lisbon.
- Friendly, M., 2005. Milestones in the History of Data Visualization: A Case Study in Statistical Historiography, in: Weihs, C., Gaul, W. (Eds.), Classification — the Ubiquitous Challenge. Springer-Verlag, Berlin/Heidelberg, pp. 34–52.  
[https://doi.org/10.1007/3-540-28084-7\\_4](https://doi.org/10.1007/3-540-28084-7_4)
- Gershon, N., Eick, S.G., Card, S., 1998. Information visualization. Interactions 5, 9–15.  
<https://doi.org/10.1145/274430.274432>
- Gershon, N., Page, W., 2001. What storytelling can do for information visualization. Commun. ACM 44, 31–37. <https://doi.org/10.1145/381641.381653>
- Grandin, F.R., 2014. A contribuição do jornalismo guiado por dados na criação de valor nas organizações jornalísticas. Universidade do Estado do Rio de Janeiro, Rio de Janeiro.
- Holovaty, A., 2006. A fundamental way newspaper sites need to change [WWW Document]. Holovaty.com. URL <http://www.holovaty.com/writing/fundamental-change/> (accessed 9.30.17).
- Howard, A.B., 2014. The Art and Science of Data-Driven Journalism. Columbia University. <https://doi.org/10.7916/D8Q531V1>
- Jensen, J.F., 2008. The concept of interactivity -- revisited: four new typologies for a new media landscape. ACM Press, p. 129.  
<https://doi.org/10.1145/1453805.1453831>
- Kirk, A., 2010. Wikileaks war logs: a triumph in “Data Journalism”? [WWW Document]. Vis. Data. URL <http://www.visualisingdata.com/2010/07/wikileaks-war-logs-a-triumph-in-data-journalism/> (accessed 11.17.17).
- Knaflic, C.N., 2015. Storytelling with data: a data visualization guide for business professionals. Wiley, Hoboken, New Jersey.
- Knight, M., 2015. Data journalism in the UK: a preliminary analysis of form and content. J. Media Pract. 16, 55–72.  
<https://doi.org/10.1080/14682753.2015.1015801>
- Krum, R., 2014. Cool infographics: effective communication with data visualization and design. Wiley, Indianapolis, IN.

- Lee, B., Riche, N.H., Isenberg, P., Carpendale, S., 2015. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Comput. Graph. Appl.* 35, 84–90. <https://doi.org/10.1109/MCG.2015.99>
- Legewie, N., 2013. An Introduction to Applied Data Analysis with Qualitative Comparative Analysis (QCA). *Forum Qual. Sozialforschung Forum Qual. Soc. Res.* 14. <http://dx.doi.org/10.17169/fqs-14.3.1961>
- Lewis, S.C., Westlund, O., 2015. Big Data and Journalism: Epistemology, expertise, economics, and ethics. *Digit. Journal.* 3, 447–466. <https://doi.org/10.1080/21670811.2014.976418>
- Lima, S., 2016. Jornalismo de Dados no Brasil: Tendências e desafios, in: *Monitoramento e Pesquisa em Mídias Sociais: metodologias, aplicações e inovações*. Uva Limão, São Paulo, pp. 279–300.
- Loosen, W., Reimer, J., De Silva-Schmidt, F., 2017. Data-driven reporting: An on-going (r)evolution? An analysis of projects nominated for the *Data Journalism Awards* 2013–2016. *Journal. Theory Pract. Crit.* 146488491773569. <https://doi.org/10.1177/1464884917735691>
- Lorenz, M., 2010. Status and Outlook for data-driven journalism.
- Lorenz, M., n.d. Mirko Lorenz [WWW Document]. Mirko Lorenz. URL [www.mirkolorenz.com](http://www.mirkolorenz.com) (accessed 1.4.18).
- Maltese, A., Svetina, D., Harsh, J., 2015. Research and Teaching: Data Visualization Literacy: Investigating Data Interpretation Along the Novice-Expert Continuum. *J. Coll. Sci. Teach.* 045. [https://doi.org/10.2505/4/jcst15\\_045\\_01\\_84](https://doi.org/10.2505/4/jcst15_045_01_84)
- Mancini, L., Vasconcellos, F., 2016. Jornalismo de Dados: conceito e categorias. *Front. - Estud. Midiáticos* 18. <https://doi.org/10.4013/fem.2016.181.07>
- Manovich, L., 2012. Trending: The Promises and the Challenges of Big Social Data, in: Gold, M.K. (Ed.), *Debates in the Digital Humanities*. University of Minnesota Press, pp. 460–475. <https://doi.org/10.5749/minnesota/9780816677948.003.0047>
- Martinho, A.P., 2013. Jornalismo de Dados: Contributo para uma Caracterização do Estado da Arte em Portugal. ISCTE - Lisbon University Institute, Lisbon.
- Mayer-Schönberger, V., Cukier, K., 2013. *Big data: a revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, Boston.
- Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army [WWW Document], n.d. . R. Collect. Trust. URL <https://www.royalcollection.org.uk/collection/1075240/notes-on-matters-affecting-the-health-efficiency-and-hospital-administration-of>
- Reilly, S., 2017. The Need to Help Journalists with Data and Information Visualization. *IEEE Comput. Graph. Appl.* 37, 8–10. <https://doi.org/10.1109/MCG.2017.32>
- Rogers, S., 2013a. History of data journalism at the Guardian [WWW Document]. The Guardian. URL <https://www.theguardian.com/news/datablog/video/2013/apr/04/history-of-data-journalism-video> (accessed 11.11.17).
- Rogers, S., 2013b. *Facts are sacred: the power of data*. Faber and Faber, Guardian Books, London.
- Rogers, S., 2012. Behind the Scenes at the Guardian Datablog, in: Bounegru, L., Gray, J., Chambers, L. (Eds.), *The Data Journalism Handbook*. O'Reilly Media, Sebastopol, CA.



- Rogers, S., 2011. Data journalism broken down: what we do to the data before you see it. Datablog.
- Rogers, S., 2010. Florence Nightingale, datajournalist: information has always been beautiful. Datablog.
- Rogers, S., 2008. Turning official figures into understandable graphics, at the press of a button. Guard. Blog.
- Rogers, S., Schwabish, J., Bowers, D., 2017. Data Journalism in 2017: the current state and the challenges facing the field today. Google News Lab.
- Sachs, J., n.d. 1786/1801: William Playfair, Statistical Graphics, and the Meaning of an Event. BRANCH Br. Represent. Ninet.-Century Hist.
- Segel, E., Heer, J., 2010. Narrative Visualization: Telling Stories with Data. IEEE Trans. Vis. Comput. Graph. 16, 1139–1148. <https://doi.org/10.1109/TVCG.2010.179>
- Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. IEEE Comput. Soc. Press, pp. 336–343. <https://doi.org/10.1109/VL.1996.545307>
- Shneiderman, B., Plaisant, C., 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. ACM Press, p. 1. <https://doi.org/10.1145/1168149.1168158>
- Snyder, J., 2017. Vernacular Visualization Practices in a Citizen Science Project. ACM Press, pp. 2097–2111. <https://doi.org/10.1145/2998181.2998239>
- Splendore, S., Di Salvo, P., Eberwein, T., Groenhardt, H., Kus, M., Porlezza, C., 2016. Educational strategies in data journalism: A comparative study of six European countries. Journal. Theory Pract. Crit. 17, 138–152. <https://doi.org/10.1177/1464884915612683>
- Stone, M.L., 2014. Big data for media. Reuters Institute for the Study of Journalism, Oxford, UK.
- The first Guardian data journalism: May 5, 1821, 2011. . Datablog.
- Träsel, M., 2013. O jornalismo guiado por dados numa perspectiva brasileira, in: Gray, J., Bounegru, L., Chambers, L. (Eds.), Manual de Jornalismo de Dados: Como Os Jornalistas Podem Usar Dados Para Melhorar Suas Reportagens. Abraji/EJC, São Paulo.
- Tufte, E.R., 1997. Visual and statistical thinking: displays of evidence for making decision, 4. print. ed. Graphics Press, Cheshire, Conn.
- Tufte, E.R., 1983. The visual display of quantitative information. Graphic Press, Cheshire, CT.
- van Wijk, J.J., 2005. The Value of Visualization, in: Proc. IEEE Conf. Visualization. pp. 79–86.
- Vasconcellos, F., Mancini, L., Bittencourt, C., 2015. Cinco categorias de Jornalismo de Dados ou uma proposta para problematizar o Jornalismo a partir de dados no Brasil.
- Veglis, A., Bratsas, C., 2017. Towards A Taxonomy of Data Journalism. J. Media Crit. 3, 109–121. <https://doi.org/10.17349/jmc117309>
- Viégas, F.B., Wattenberg, M., 2008. Tag clouds and the case for vernacular visualization. interactions 15, 49. <https://doi.org/10.1145/1374489.1374501>
- Weber, W., Rall, H., 2012. Data Visualization in Online Journalism and Its Implications for the Production Process. IEEE, pp. 349–356. <https://doi.org/10.1109/IV.2012.65>

- Yau, N., 2017. This is an American Workday, By Occupation [WWW Document]. FlowingData. URL <https://flowingdata.com/2017/05/17/american-workday/> (accessed 12.18.17).
- Yi, J.S., Kang, Y. ah, Stasko, J., 2007. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Trans. Vis. Comput. Graph.* 13, 1224–1231. <https://doi.org/10.1109/TVCG.2007.70515>
- Yin, R.K., 1994. Case study research: design and methods, 2nd ed. ed, Applied social research methods series. Sage Publications, Thousand Oaks.
- Young, M.L., Hermida, A., Fulda, J., 2017. What Makes for Great Data Journalism? *Journal. Pract.* <https://doi.org/10.1080/17512786.2016.1270171>
- Zanchelli, M., Crucianelli, S., 2013. Integrating Data Journalism into Newsrooms. International Center for Journalists and Knight International Journalism Fellowships.
- Zikopoulos, P., Eaton, C., deRoss, D., Deutsch, T., Lapis, G., 2012. Understanding big data: analytics for enterprise class Hadoop and streaming data. McGraw-Hill, New York.

**Figure 59 - Manchester Guardian, 5th May 1821**

[illegible]

**BANKRUPTS.**  
Gazette—Saturday, April 22.  
CLAUDE Thomas, of Gainsborough, in the  
of Lincoln, docket at the County Court, at Lincoln.

11, at the Punch Bowl, Nottingham: At  
Payne, Nottingham.

[illegible][illegible]

ALLISON George, of Bishop Worcester, county of Durham, says mately 4 c. (see also *W. A. B. 1896*)  
 BAYLY William Henry, of Cheltenham, in the county of Gloucester, banker, d. c. May 11, 1896, aged 66. He had a wife and 10 children.  
 DENMAISE William, of Buxley, in the parish of Otley, in the county of York, worked as a labourer, d. c. Nov. 16, New Inn, Bradford. Mr. Crookall, Bradford.  
 FRANKIE Richard the Head, of Newark upon Trent, in the county of Nottingham, d. c. Aug. 9, 1896, at Newark's Head Inn, Newark. Messrs. Tallents and Secord, Newark.  
 GOSWELL Thomas, of the county of Lancaster, upholsterer, d. c. May 29, 1896, 22, White Hart Inn, Freetown, Lancashire. Trough, Freetown.  
 JERRY John, of Kirtley, in the county of Stafford, d. c. May 16, 1896, at Kirtley, Stafford. Messrs. Crookall and Hones, Ipswich.  
 SHAW David, of the county of the parish of Kirtley, in the county of Worcester, and the village of Hinton, in the parish of Bishop, in the county of Worcester, d. c. May 16, 1896, at Kirtley, Worcester. Messrs. Crookall and Hones, Ipswich.

[illegible]

**MANCHESTER;**  
Printed and published by J. GARNETT, No. 11, Market-street; where orders and advertisements will be received; and also by all Booksellers.

Figure 60 - Explanation of how to read the line chart that shows the magnitude of cholera deaths in New York ("New-York Daily Tribune., September 29, 1849, Image 1," n.d.)

DRAWN AND ENGRAVED EXPRESSLY FOR THE NEW-YORK TRIBUNE.

The above Diagram, or *Graphic Table*, (for which we are indebted to Professor GILLSPY of Union College,) represents to the eye, in a very striking manner, the rise, progress, and decline of the Cholera, and other diseases in our City, during the last four months.

Each half-inch along the bottom line represents a week. The dates are placed under each. At the end of each half-inch, or week, are upright dotted lines, whose various lengths indicate the number of deaths by Cholera, and other causes, during that week; each inch on these upright lines corresponding to 500 deaths. The numbers are placed at the top of each. The zig-zag lines, which join the ends of these lines, show, by their upward or downward slopes, whether the deaths during those weeks have increased or decreased, rapidly or slowly.

Some curious circumstances are apparent, on an inspection of this figure. We see that during the first two weeks of the Cholera, while it was increasing, as shown by the line sloping upward, the total deaths were decreasing, as shown by the line sloping downward. This was probably caused by the great care in diet, &c. then practiced, on the first alarm. In the week ending July 7, while the Cholera line goes up the "Total deaths" line unaccountably goes down. Thence both lines go on ascending, and the corresponding number of deaths increasing, till in the week ending July 21, they reach their culminating point. Thence there is a constant and quite regular descent, excepting a sudden fall, from Aug. 4 to 11. This may have arisen from the diminution of population, consequent upon the custom of many persons to leave the city at the beginning of that month. The descent keeps on, till it brings up to our present normal condition.

If the average temperature, moisture, electrical state, &c. during these weeks were represented in the same manner, and added to this diagram, their comparison would show at a glance whether there has been any connection between them.

Figure 61 - Overview of “*Música Muito Popular Brasileira*” (Folha): due to its extension, the story was divided in 5 columns

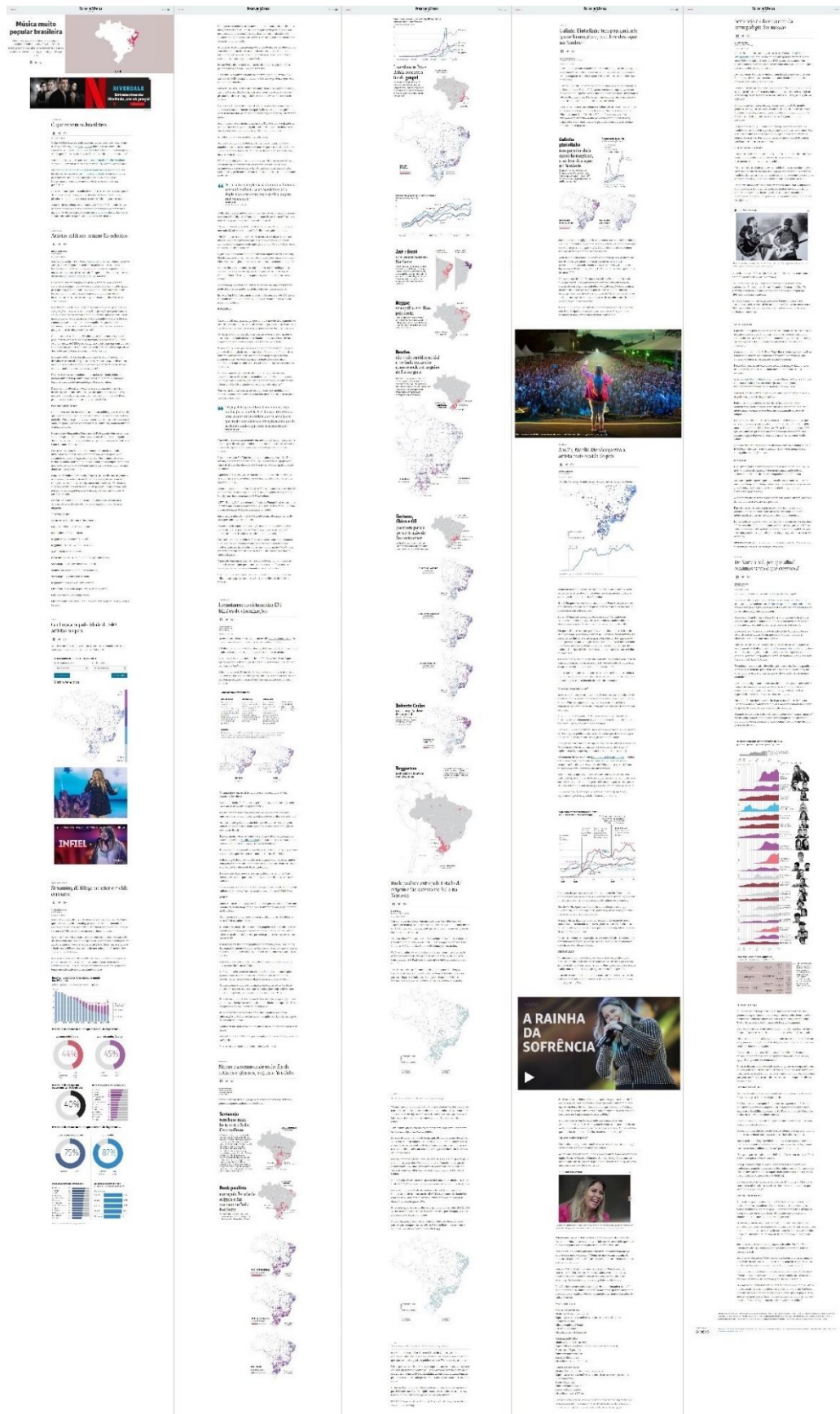
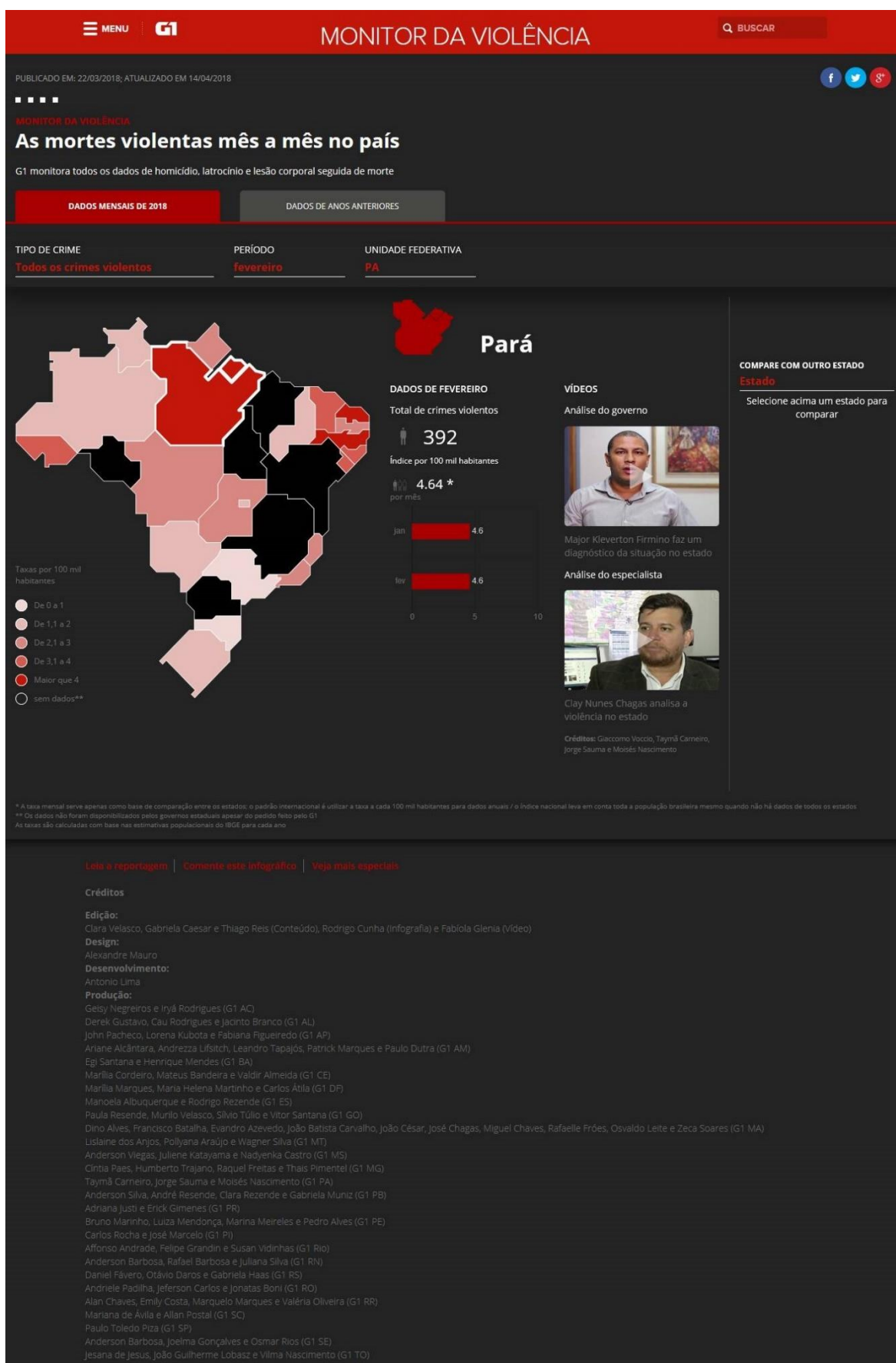




Figure 62 - Overview of “*Barcelona, capital catalã, teve maior proporção de votos contrários à separação*” - O Globo



Figure 63 - Overview of “As mortes violentas mês a mês no país” (G1)



**Figure 64 - Overview of “Como sua cidade gasta dinheiro? Descubra e compare com outras” (Nexo):** each column represents possible configurations, according to the filter applied. The first shows information about a city. The second is a comparison between two cities.





*the story was divided in 3 columns*



Um levantamento inédito feito pelo The Intercept Brasil, com base em informações obtidas com exclusividade de **Diágora Demência**, mostra que, das 475 ligações anônimas que o serviço recebeu em 2016 e 2017 – referente às atividades de traficantes e paramilitares na capital –, 65% delas denunciam milicianos. Como não há dados sistematizados pelo governo sobre o avanço das milícias no Rio, o volume de denúncias – analisadas por palavras-chave – são o mais forte indicativo dos caminhos de crime organizado.

**Territórios**

Quem teve mais ocorrências em cidades do Rio de Janeiro no Único Destino, **favelas** ou **vilas**? No gráfico abaixo, compare a participação de cada organização criminoso em relação a 17,8 mil denúncias recebidas pelo saraio entre 2016 e 2017.

**Excesso uma cidade: CACHOEIRAS DE MAÍ**

2016	2017
1.674,4	1.947,4

Nascidos no vício de poder público em meados dos anos 2000 sob a aura de proteção das comunidades contra o tráfico, os grupos de *"autodefesas comunitárias"* foram apoiados e encorajados por governantes, cresceram e se tornaram o que vemos hoje: *polícias e empregadas – concorrentes* que batem de frente com as forças de traficantes pelo controle de uma mina de ouro, as "prostitutas" dos bairros. Tudo isso ao custo de muito sangue. Os "segurancas comunitários" se tornaram, também eles, senhores da vida e da morte.

### Milícia is the new tráfico

"Na esquina da Rua X com Rua Y, no Morro de Abacateiro, localiza-se um comércio chamado W, onde podem ser vistos, nos finais de semana, as duas filhas do miliciano morto em 2014, de vulgar *Assul*, as quais estão tentando implantar o tráfico de drogas na localidade com o apoio de traficantes de Morro X". Era fim da manhã de 17 de março de 2017 quando o Diágo Denadim recebeu a informação de que milicianos

Disque Brindado RJ

Notas: este gráfico mostra os resultados para o mês inteiro em que ocorreram os parâmetros de controle de 2002 a 2006, e não 2007.

DISC. J. AMERIC.	DISC. J. NOROCC.	DISC. J. CARIBE	DISC. J. ASIA/PAC.	DISC. J. EUROPA	DISC. J. OCEANIA
100	100	100	100	100	100
90	90	90	90	90	90
80	80	80	80	80	80
70	70	70	70	70	70
60	60	60	60	60	60
50	50	50	50	50	50
40	40	40	40	40	40
30	30	30	30	30	30
20	20	20	20	20	20
10	10	10	10	10	10
0	0	0	0	0	0

Tema	Publicaciones
Tema 1	4,800
Tema 2	3,200
Tema 1	1,800
Tema 2	1,200
Tema 1	800
Tema 2	600
Tema 1	400
Tema 2	200
Tema 1	100
Tema 2	50

Fuente: Datos propios

Quarenta quilômetros longe dali e um ano antes, na zona norte de Rio, a única loja não era diferente. "Na avenida citada, perfumo a um mercado, localiza-se o bar, onde freqüentemente, na parte da tarde, pode ser encontrado o militante que atua na área de jogo de bicho, estende de comerciantes e tráfico de entorpecentes. O citado costuma andar sempre, acompanhado de outros militantes. Ele usa sempre

## Minha Casa, Minha Milícia

Da milicinas mancharam um fardo imenso

O policial faz vista grossa no momento da invasão, se aumentando do local. Depois que a munição se instala, o policiamento retorna, desta vez para impedir o retorno dos traficantes. Isso é um fenômeno que vem de dentro do poder: diz o então comandante do Tóp, coronel Mário Sérgio de Brito Duarte, ainda em 2004. Para ele, a ocupação desses grupos se faz possível com apoio da população e a participação de parcela das unidades policiais dessas regiões.

**Disque de entrada no**  
Número de pessoas que foram vacinadas a milhões e Pacotes  
comprados em 2015 e 2017

O gráfico de linhas compara o número de pessoas vacinadas (eixo Y, 0 a 100 milhões) e os pacotes comprados (eixo X, Jan a Dez) em 2015 e 2017. A linha verde representa as vacinações em 2015, a linha azul as vacinações em 2017, a linha tracejada verde os pacotes comprados em 2015 e a linha tracejada azul os pacotes comprados em 2017. As vacinações em 2017 foram significativamente maiores do que em 2015, especialmente no segundo semestre. Os pacotes comprados em 2017 também foram maiores do que em 2015, com um pico em setembro.

Mês	Vacinas 2015 (Milhões)	Vacinas 2017 (Milhões)	Pacotes 2015	Pacotes 2017
Jan	15	15	15	15
Fev	10	10	10	10
Mar	10	10	10	10
Abr	10	10	10	10
Mai	10	10	10	10
Jun	10	10	10	10
Jul	10	10	10	10
Ago	10	10	10	10
Sep	10	10	10	10
Out	10	10	10	10
Nov	10	10	10	10
Dez	10	10	10	10

Nessa época, uma favela dominada pelo tráfico era tomada por uma milícia a cada 12 dias. Coordenadas por agentes de segurança pública, políticos e líderes comunitários, como apenhou **relatório elaborado pelo Gabinete Militar da prefeitura de Rio**, era o início de uma época de ouro.

**Desastres nos Estados**  
Desastres por tipo de  
região: [www.crisisdata.org](http://www.crisisdata.org)

Este poder tem forte impacto na Salada Fluminense, região com histórico de violência há décadas. Com as UPPs concentradas na cidade de Rio, os traficantes migraram para

[illegible]

para 22,5 apenas dois anos depois, segundo o Instituto de Segurança Pública. Na Baixada, em 2015, a taxa era de 45,4 para cada 100 mil habitantes. Em 2017, atingiu 60,6. E esse morce é um instrumento determinado. Em 2016 – Gêmeos dados disponíveis – 79% das vítimas de mortes intencionais na Baixada foram por armas de fogo. Número que ultrapassa a média do estado, de 74%.

Os grupos paramilitares chamaram aten-

A vereadora Marielle Franco — encastada com o motorista Anderson Pedro no dia 14 de março — trabalhou exatamente nesta CPI, com o deputado Márcio Prestes. Com o fim dos trabalhos e uma festa inicial, aparentemente, pouco ou nada mudou no modo como o poder público atua. Enquanto milicianos voltaram a frequentar os palácios de poder, Prestes, ameaçado de

more, segue ali hoje com cunha política.

El Marielle foi morta, ao que tudo indica, por esse mesmo poder.

"A repressão estatal não consegue desarticular as milícias, apenas entrafugá-las. A evolução mais marcante da operação das milícias foi respeito à crescente discreção e sigilo das suas atividades, que contrastam com a ostentação de anos anteriores e que dificultaram as investigações e as prisões dos seus membros", explica Thais Duarte. "No saguão" do 11, inclusive, a operação mais silenciosa pelas autoridades de

**Ano eleitoral. Ano próspero.**

Carminha **estava grávida** quando recebeu a notícia de que tinha sido elctta. Nas eleições, saiu da cadeia e foi **recebida calorosamente** ainda no aeroporto. Chegou a ser empolgada, mas teve o **mandado cassado** seis meses depois por irregularidades na prestação de contas da campanha. Calu em uma desgraça que **podeu durar**. Depois de ser **rejeitada** até pelos próprios pares, nas últimas eleições ela votou a cara no sol mais uma vez. A ex-veradeira desafiou Freitas e agitou seu nomecanto à Prefeitura de Rio, e pastor Manoel Crivella, que agradeceu: **"tudo aqui é importante"**.

A gente percebe que estamos em ano eleitoral quando tudo isso começa", diz o ex-lder do Diáque Domênica, Ivoa Borges, que está em contato com o Tribunal Regional Eleitoral e o Ministério Público para repasse de informações. "Isso geralmente fica mais intenso até as convenções, mas continua até o voto".

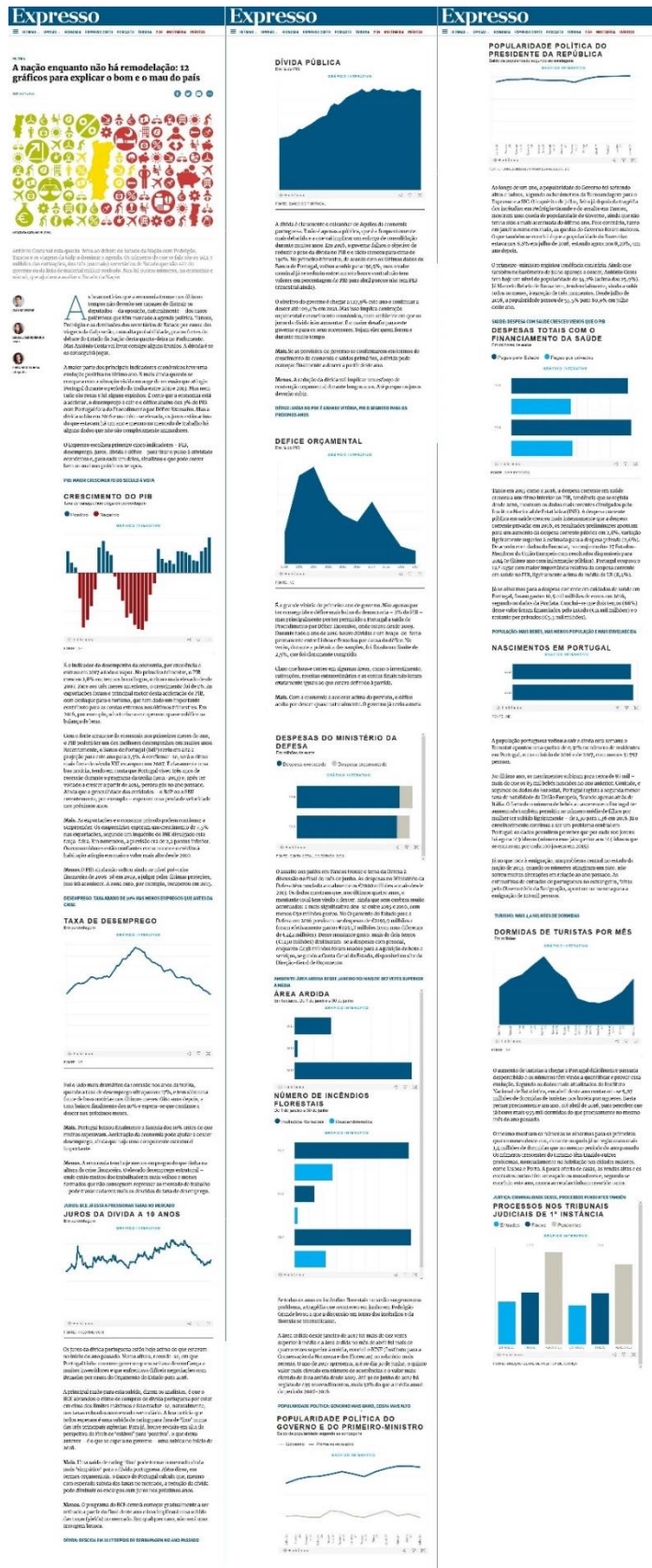
A CPI das Milícias provocou a prisão de alguns vereadores e deputados, mas a influência dos milicianos nas eleições não foi abalada. Só no último ano eleitoral, o Diáque Domênica

Novos territórios, aos ditos, pas para crescer. Suspeitas de praticar **doenças de eliminação política**, as milícias seguem atuando sem serem incomodadas, mesmo sob a intervenção militar federal que já dura mais de um mês. As operações do Exército ignoram as áreas dominadas por paramilitares. O interventor chegou a dizer que **agita contra a corrupção** na polícia, mas as visitas feitas em Estâncias focaram apenas nas **condições de trabalho**.

"Morei na Praça Seix por 24 anos e nunca vi a situação como está hoje. Não que não houveresse violência, mas eram muito raras. De um cinco anos por ex, elas passaram de intensidade e frequência", explica um comendador da local que pediu anonimato já que hoje vive em um bairro vizinho. Frequentemente, corpos são deixados em cima da via, geralmente em áreas de estacionamento ou perto de lojas. Não são

Esses cenários não são novos e os precedentes deixam a dica de que pouco ou nada vai mudar. O interventor federal no Estado foi o coordenador da segurança na Olimpíada. Neste mesmo período, os agentes da Força Nacional que participaram do esquema de segurança dos Jogos tiveram de se submeter à ordem da milícia na Zona Oeste do Rio.

Figure 66 - Overview of “[A nação enquanto não há remodelação: 12 gráficos para explicar o bom e o mau do país](#)” (Expresso): due to its extension, the story was divided in 3 columns



**Figure 67 - Overview of “*Em 20 anos, o que mudou em Portugal?*” (Público): due to its extension, the story was divided in 3 columns**

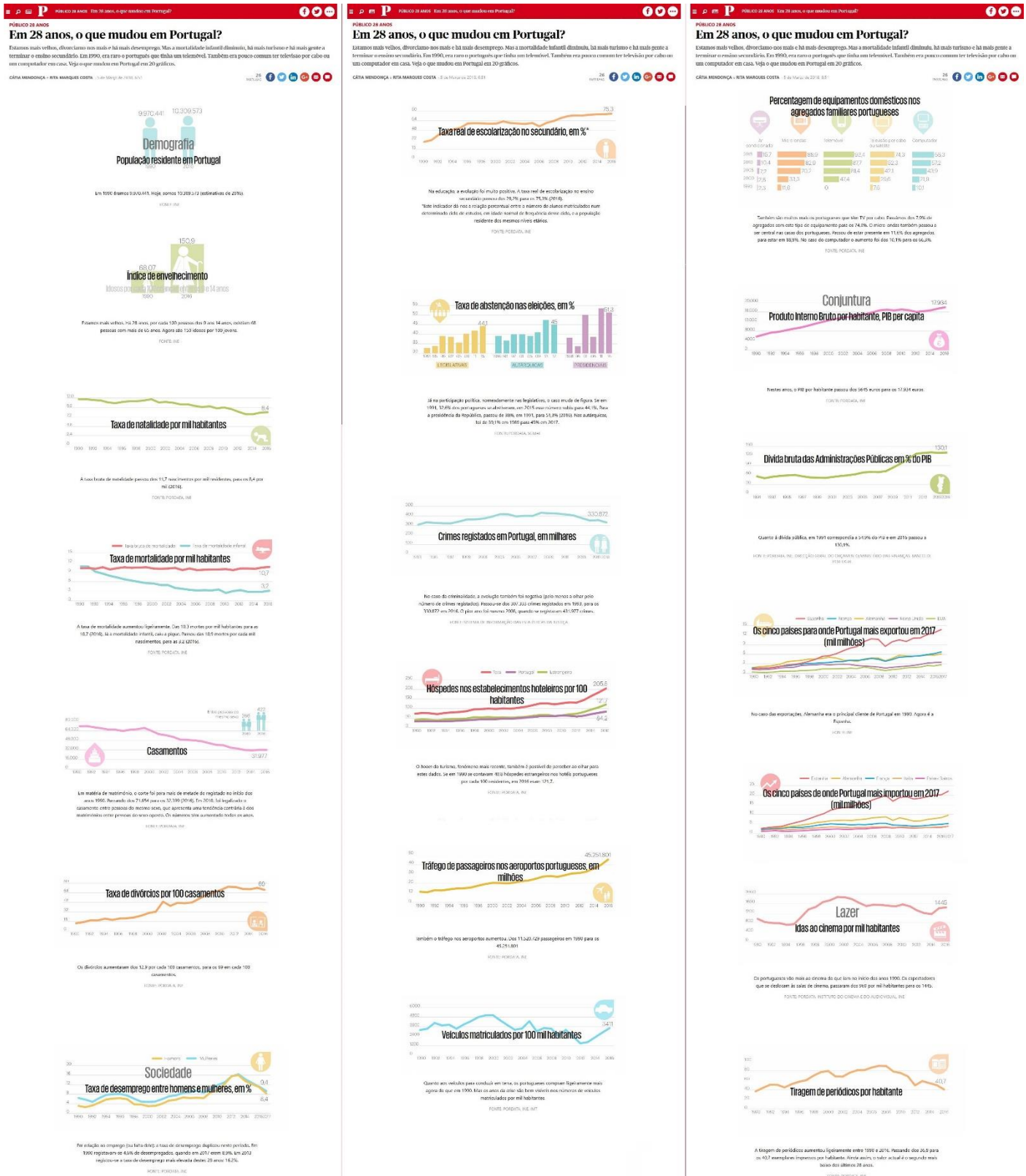


Figure 68 - Overview of “[Mapa interativo. Sabe o significado literal do nome dos países?](#)” (Observador)





**Figure 69** - Overview of "Benefícios a juízes nos Estados têm disparidades e distorções" (Folha): due to its extension, the story was divided in 2 columns



Figure 70 - Overview of ["Com extras, 71% dos juizes do país recebem acima do teto de R\\$ 33 mil"](#) (O Globo)



**Figure 71 - Overview of “TJs gastam R\$ 890 mi com ‘penduricalhos’ para juízes” (Estadão)**

[illegible]

Figure 72 - Overview of “Quanto ganham desembargadores e juizes no Brasil” (Nexo): due to its extension, the story was divided in 2 columns

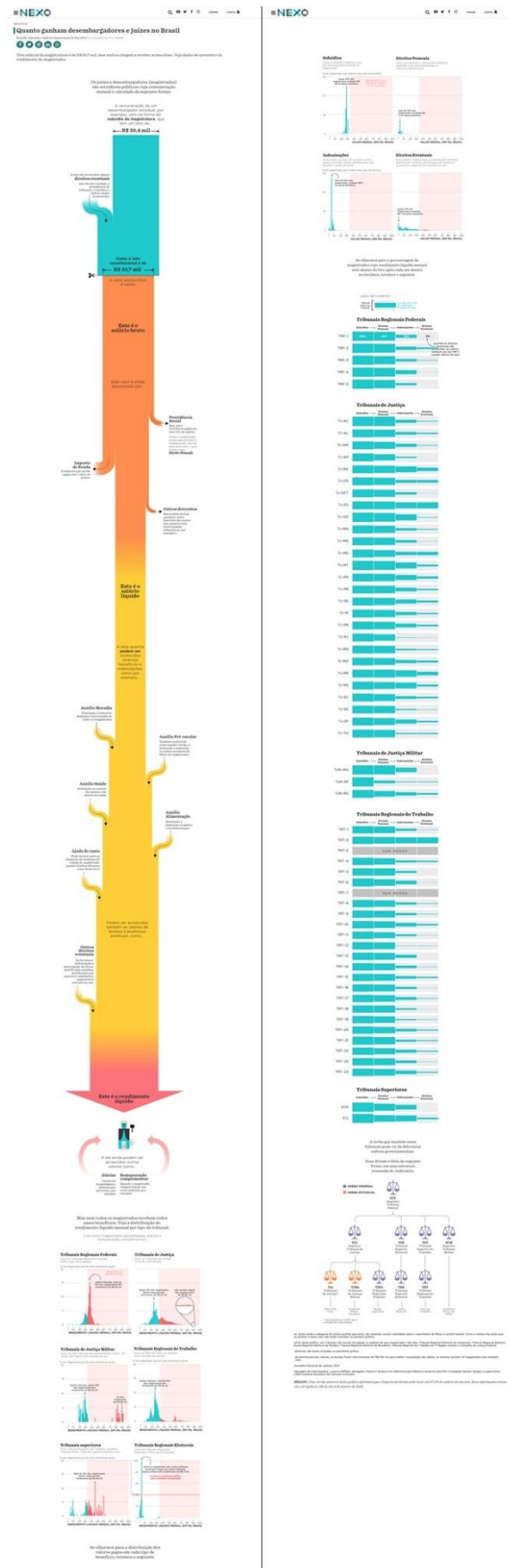
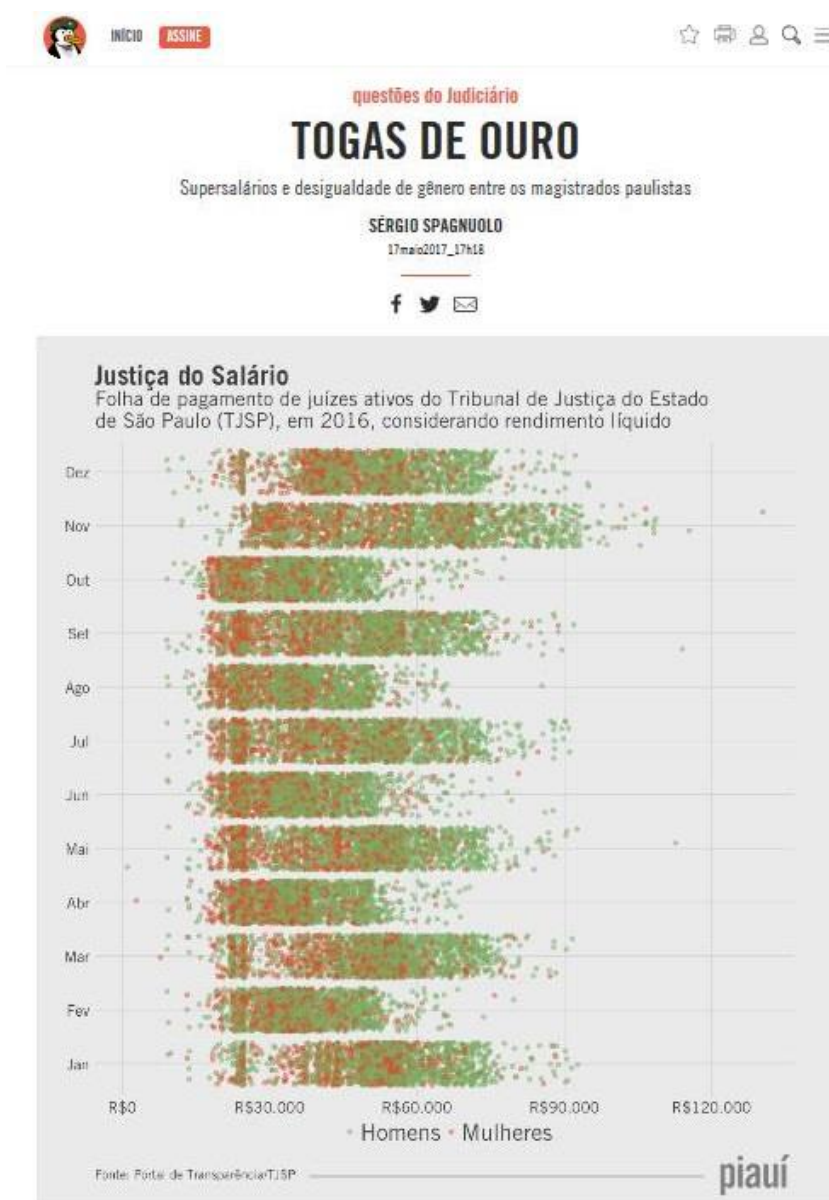




Figure 73 - Overview of “*Togas de ouro*” (Volt Data Lab for Piauí Magazine)



**E**m 2016, os 2 591 juizes e desembargadores do Tribunal de Justiça de São Paulo ganharam, em média, 42 mil reais líquidos por mês. Ainda que o teto salarial da magistratura paulista fosse de 30,4 mil reais, muitos deles conseguiram ultrapassar o limite legal de vencimentos porque somaram vantagens extras – como subsídios, indenizações e gratificações –, que não entram na conta.

A maior bolada mensal foi recebida pelo desembargador José Carlos Gonçalves Xavier de Aquino. Apenas em novembro, ele embolsou 130 mil reais.

A desigualdade salarial entre gêneros também é visível na Corte que, até hoje, jamais foi presidida por uma mulher. Juizes homens ganham em média 44 mil reais por mês – 13,5% a mais do que as mulheres (38 mil reais mensais). No caso de desembargadores, a diferença é maior: homens faturaram 55 mil reais por mês, 16,5% a mais do que as mulheres, que ganham em média 46 mil reais.



SÉRGIO SPAGNUOLO

TAGS: DESIGUALDADE DE GÊNERO MAGISTRADO SALÁRIOS SÃO PAULO SÉRGIO SPAGNUOLO

Figure 74 - Overview of [“A estrada da morte: onde o fogo matou – e onde ainda lava \(mapa interativo\)”](#) (Expresso)

SOCIEDADE

## A estrada da morte: onde o fogo matou – e onde ainda lava (mapa interativo)

18.06.2017 às 11h09



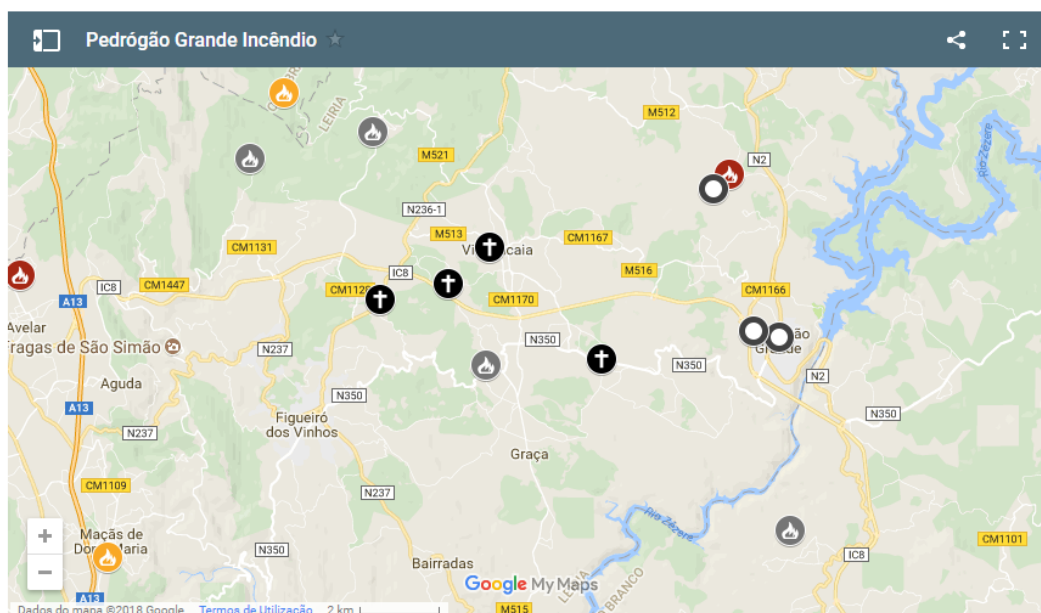
O incêndio de Pedrógão Grande, já considerado o mais violento de sempre no país, provocou até ao momento 62 mortos, dos quais 47 foram encontrados na estrada nacional 236, que faz a ligação com o IC8 e que não estava cortada. Das 62 vítimas mortais, 30 estavam dentro das suas viaturas e 17 fora delas ou à beira da estrada



SOFIA MIGUEL ROSA

● Incêndios ativos ● Vítimas mortais ○ Apoio às vítimas

MAPA INTERATIVO EM ATUALIZAÇÃO



FONTE: PROTEÇÃO CIVIL

### Palavras-chave

TRAGÉDIA EM PEDRÓGÃO

Figure 75 - Overview of “As freguesias nas listas de prioridade de limpeza de terrenos” (Jornal de Notícias)

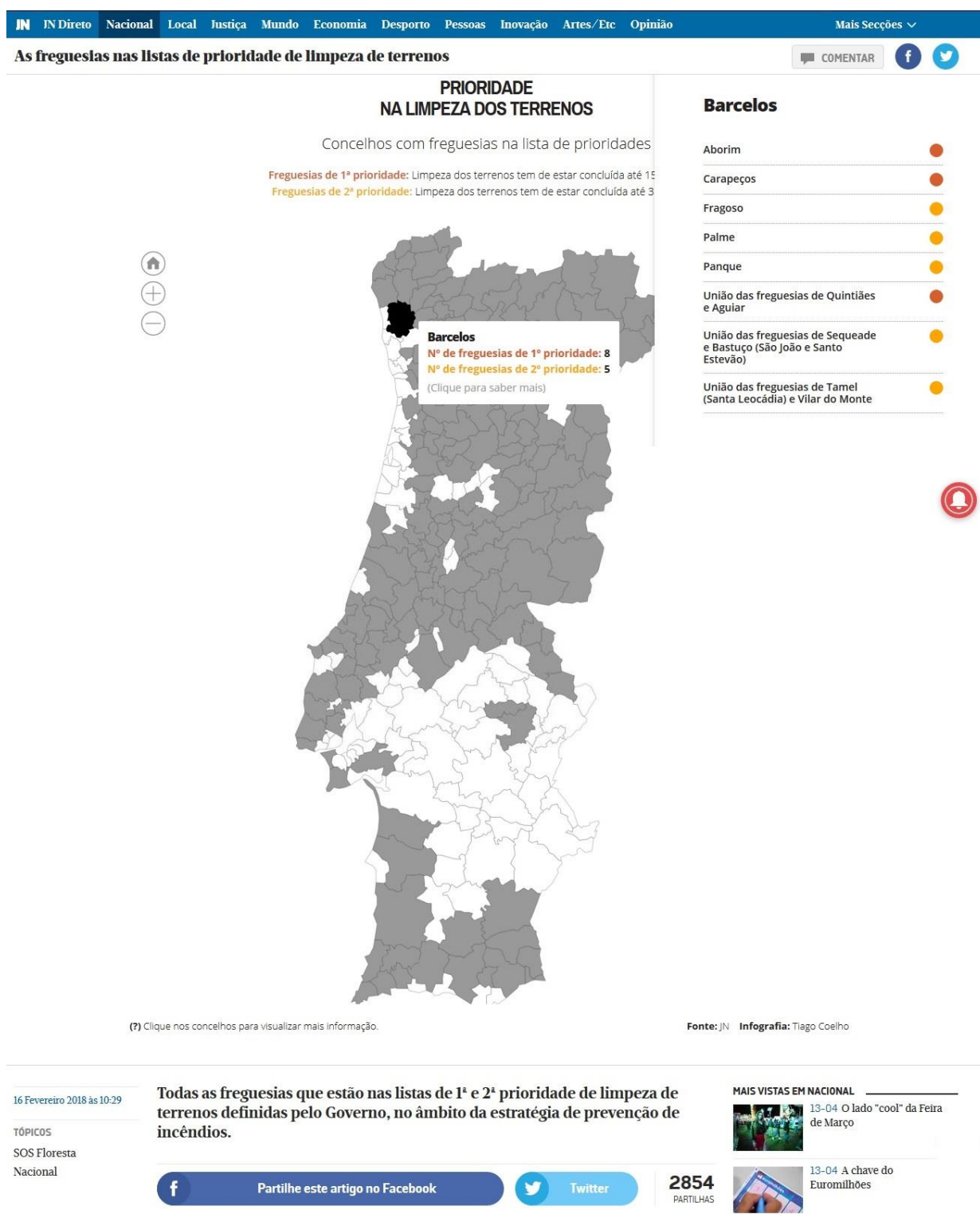


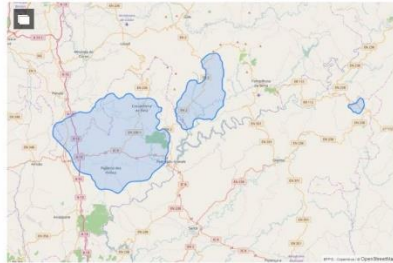
Figure 76 – Overview of “Incêndio de Pedrógão Grande é o maior de sempre em Portugal” (Público)

## INCÊNDIOS FLORESTAIS Incêndio de Pedrógão Grande é o maior de sempre em Portugal

Chamas já consumiram mais de 30 mil hectares de floresta, superando os números negros dos piores incêndios de 2003 e 2012.

ROMANA BORJA-SANTOS • SOFIA RODRIGUES • 19 de Junho de 2017, 17:47 (atualizado a 19 de Junho de 2017, 18:43)

5071  
partilhados



O incêndio na zona de Pedrógão Grande já consumiu mais de 30 mil hectares de floresta desde sábado passado, segundo dados do Sistema Europeu de Informação de Fogos Florestais (conhecido pela sigla EFFIS). Este valor ultrapassa largamente os números atingidos nos grandes incêndios de 2012 em Tavira e de 2003 na Chamusca — os dois maiores fogos até agora em Portugal.

No distrito de Leiria, entre Figueiró dos Vinhos, Castanheira de Pera e Aguda já arderam 25.969 hectares desde o passado sábado. Já no concelho de Pampilhosa da Serra o fogo consumiu outros 7310 hectares, a que acrescem outros 481 hectares no concelho de Oleiros, distrito de Castelo Branco, segundo o EFFIS. Neste distrito, a área ardida foi contabilizada desde domingo.



Ao minuto: “Podemos estar a assistir a uma evolução favorável”, diz Marçalo. 18.644 mortos confirmados

Estes valores colocam a tragédia de sábado como a mais negra de sempre no país em termos de incêndios. Não só em número de vítimas, mas também de área ardida. O fogo de Cachopo, em Tavira, em Julho de 2012, passa agora a ser o segundo maior, com 24.843 hectares de área ardida, seguido pelo de Ulme, na Chamusca, que em Agosto de 2003 destruiu 22.190 hectares.

No ano passado, os incêndios florestais mais graves aconteceram nos concelhos de Arouca e São Pedro do Sul, onde o fogo consumiu no total cerca de 29 mil hectares de terreno. No período entre Janeiro e Outubro de 2016, a área ardida contabilizada foi de 160.490 hectares, um recorde nos últimos dez anos, de acordo com o relatório do Instituto da Conservação da Natureza e das Florestas (ICNF) e que só tem dados até ao último trimestre.

Já neste ano, até dia 15 de Junho, foram registados 5760 incêndios em Portugal, que consumiram 15.184 hectares. Os dados mais recentes do ICNF ainda não incluem o trágico incêndio do último fim de semana, mas mostram que o cenário já era negativo e que vai piorar. Nestes primeiros seis meses do ano, a área ardida já representava mais 60% do que a média da área ardida entre 2007 e 2016, de 9636 hectares.

### Os cinco maiores incêndios em Portugal

Freguesia	Concelho	Área ardida (ha)	Início	Extinção
Pedrógão Grande	Pedrógão Grande	25.969	17-06-2017	
Cachopo	Tavira	24.843	18-07-2012	20-07-2012
Ulme	Chamusca	22.190	02-08-2003	07-08-2003
Intense	Arouca	21.910	08-08-2016	13-08-2016
Beirer	Gondal	20.088	01-08-2003	06-08-2003

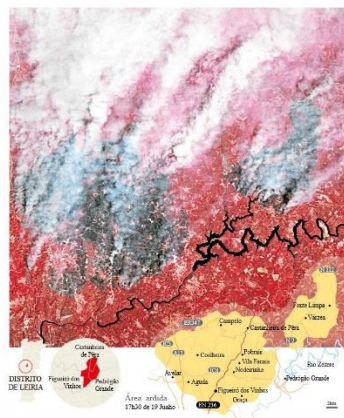


\* 1 de Janeiro a 15 de Outubro; \*\* 1 de Janeiro a 15 de Junho

Fonte: 2007-2016, ICNF; 2017, ICNF - Unidade Europeia, Sistema Europeu de Informação de Fogos Florestais

Tanto o número de ocorrências como a área ardida estão acima da média registada entre 2007 e 2016 para o mesmo período de 1 de Janeiro a 15 de Junho. Em termos de área ardida, o valor só tinha sido ultrapassado em 2009, 2012 e 2015, mostram os dados enviados ao PÚBLICO. O número de ocorrências é o terceiro mais elevado neste período.

“Comparado os valores do ano de 2017 com o histórico dos últimos dez anos destaca-se que se registaram mais 32% de ocorrências e mais 58% de área ardida relativamente à média verificada no decénio 2007-2016”, explica o ICNF, que destaca também o elevado número de reacendimentos.



Fonte: Direção-Geral do Terras e dos Recursos Naturais

PÚBLICO





Figure 77 - Overview of ["Doze anos de incêndios"](#) (Público)



Pegámos na lista de incêndios ocorridos entre 2001 e 2012 em Portugal continental, disponibilizada pelo ICNF, e construímos uma visualização que permite saber em que anos houve mais incêndios, em que municípios ocorreram e a que horas foram dados os alertas para esses fogos.

2001 - 2012

Arraste para ver um período diferente

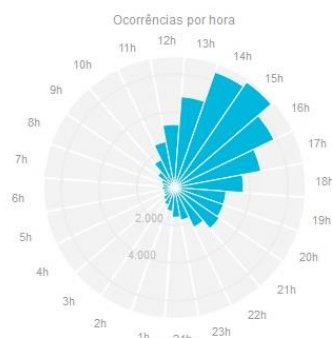
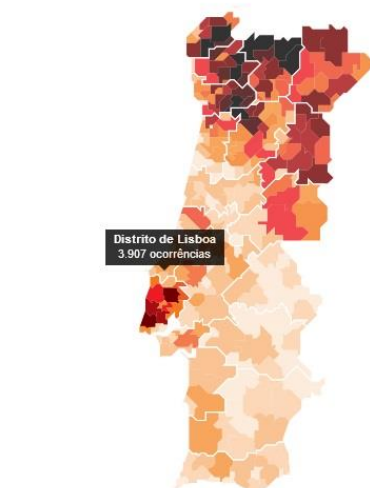
TODO O PAÍS

61.070

Total de ocorrências



Ocorrências por ano



**SOBRE OS DADOS:** No âmbito de um projecto de investigação em jornalismo computacional (REACTION), compilámos numa só lista os registos anuais dos incêndios entre 2001 e 2012, disponibilizados no site do ICNF (Instituto de Conservação da Natureza e Florestas). Retirámos não só a informação sobre a localização geográfica (por município) como também as horas e datas de alerta de cada um dos incêndios, para construção da base de dados. Cada incêndio corresponde a um fogo em que ardeu pelo menos um hectare. Fora da lista ficam as queimadas, os fogachos e os falsos alarmes e os reacendimentos dos quais tenha resultado uma área ardida inferior a um hectare.

#### COORDENAÇÃO

Joaquim Guerreiro  
Raquel Albuquerque

#### TEXTOS

Ana Fernandes  
Raquel Albuquerque

#### VÍDEO

Joana Bourgard

#### FOTOGRAFIA

Daniel Rocha  
Enric Vives-Rubio  
Joana Bourgard  
Rui Gaudêncio

#### INFOGRAFIA E DESIGN

Cátia Mendonça  
Célia Rodrigues  
José Alves  
Andrea Espadinha  
Dinis Correia

#### MÚSICA

"Hidden Blues" por Pixt  
"Exurb" por Gurdonark  
"Morning Dream" por Gurdonark  
"Ethereal Space" por Snowflake

Figure 78 - Overview of “[Veja o mapa interativo dos incêndios no centro do país](#)” (Observador)

INCÊNDIOS

# Veja o mapa interativo dos incêndios no centro do país

27/7/2017, 0:10

Cerca de 3500 bombeiros combatiam esta quinta-feira os incêndios em todo o país, com os esforços concentrados na região centro.

Partilhe

João Pereira

Autor

Observador

Email

Mais sobre

CASTELO BRANCO

INCÊNDIOS

PORTALEGRE

SANTARÉM

PAÍS

ACIDENTES E DESASTRES

SOCIEDADE

As chamas consomem desde domingo o concelho da Sertã, distrito de Castelo Branco, e já se alastraram a Proença-a-Nova, no mesmo distrito, e a Mação, no distrito de Santarém. Durante esta quinta-feira, o incêndio foi dado como “dominado” pela Proteção Civil.

Mais de 3000 bombeiros combatem as chamas de norte a sul mas os principais esforços estão concentrados nesta região. Acompanhe o mapa interativo do Observador dos incêndios no centro do país. Os dados apresentados são retirados da página da Proteção Civil e atualizados assim que possível.

Acompanhe o nosso [liveblog](#) para mais desenvolvimentos.

Incêndio em Nisa é o único que continua ativo - como aconteceu

O fogo em Nisa, em Portalegre, é o único que continua ativo nesta manhã de sexta-feira. Os incêndios dos últimos dias fizeram, pelo menos, 55 feridos ligeiros.

Por Observador

Ler artigo

Partilhar artigo

OBSERVADOR

PONTO DE SITUAÇÃO

INCÊNDIOS NO CENTRO DO PAÍS 20:30 27 de julho

PASSE COM O RATO SOBRE O MAPA PARA OBTER MAIS INFORMAÇÕES

Fuente: Página Oficial da Proteção Civil

Os dados disponibilizados pela Proteção Civil são atualizados de 10 em 10 minutos. Os dados apresentados neste mapa podem não ser os mais atuais. Por "operacionais", entende-se Bombeiros, Força Especial de Bombeiros, PSP, Forças Armadas, INEM, Equipas Sapadoras Florestais, GNR, Grupos Intervenção de Proteção e Socorro

790 operacionais

Número de operacionais no local

225 viaturas

Número de viaturas no local

13 meios aéreos

Número de meios aéreos (aviões e helicópteros) no local

Partilhar

OBSERVADOR

XX

Figure 79 - Overview of "Área ardida em Portugal" - Frames (for Observador)

OBSERVADOR

INCÊNDIOS

Ministério do Ambiente investe milhões nos municípios afetados pelos fogos

21/4/2018, 15:49

O ministro do Ambiente anunciou este sábado, em Alijó, um investimento de 16,2 milhões de euros para intervir em 1360 quilómetros de rede hidrográfica de 57 concelhos afetados pelos incêndios em 2017.

Partilhe

Capoulas Santos anunciou um investimento de mais de 16 milhões de euros nos municípios afetados  
ANTÓNIO COTRIM/LUSA

Autor

Agência Lusa

Mais sobre

AMBIENTE

INCÊNDIOS

CIÊNCIA

ACIDENTES E DESASTRES

SOCIEDADE

ÁREA ARDIDA EM PORTUGAL

Povoamentos e matos ardidos por ano, entre 1 de janeiro e 31 de outubro, em milhares de hectares

Fonte: ICNF a 03/01/2018

Notas

Sugestões?

O governante presidiu este sábado, em Alijó, distrito de Vila Real, à assinatura de protocolos com 18 municípios do Norte, que vão receber dois milhões de euros para investir em "projetos urgentes e inadiáveis" de regularização fluvial.

Depois, até à primeira semana de maio, serão assinados protocolos com autarquias do Centro do país, que vão receber quatro milhões de euros, e do Tejo, que vão receber seis milhões de euros. As verbas são disponibilizadas ao abrigo do Fundo Ambiental e por intermédio da Agência Portuguesa do Ambiente.

Trata-se claramente de prevenir e de criar um território mais resistente e trata-se mesmo de melhorar, introduzindo aqui técnicas de engenharia natural que já têm alguma robustez científica em Portugal, mas que não são tão comuns como isso", afirmou.

João Pedro Matos Fernandes salientou que se vão fazer obras "sem betão". "A rede hidrográfica é fundamental na estruturação do território (...) O conjunto de ecossistemas que existem à volta desta mesma rede hidrográfica é absolutamente fundamental para estruturar o território e dar-lhe resiliência. Dito de outra forma, uma rede hidrográfica limpa, que tem nas suas margens a vegetação que é suposto ter, é mais resistente ao fogo", referiu.

Para João Pedro Matos Fernandes, este é um "trabalho fundamental" para recuperar áreas que foram afetadas, como também algumas infraestruturas, como praias fluviais, pequenas pontes e açudes. E é, acrescentou, "sobretudo muito importante como tarefa de prevenção estrutural contra incêndios".

As intervenções e as verbas aplicadas em cada município são de dimensões diferentes, sendo que, por exemplo, para Oleiros, foi afetado um milhão de euros, e visam limpar as ribeiras para garantir as condições de escoamento e realizar intervenções em margens e nos taludes para diminuir o risco de erosão.

xxi

## **APPENDIX 2: INTERVIEWS TRANSCRIPTIONS (IN PORTUGUESE)**

### **I. ENTREVISTA COM FÁBIO TAKAHASHI – FOLHA DE S. PAULO**

#### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Fábio Takahashi, 37 anos, coordenador do núcleo de inteligência da Folha de S. Paulo. Sou formado na Metodista, em São Paulo. Estou na Folha de São Paulo desde 2003. Até 2015, fui repórter de Educação no caderno Cotidiano. Por um breve período, coordenei um curso de trainees da Folha. Em 2015, saí para fazer uma fellowship na Columbia University e, na volta, vim direto para o então recém-criado núcleo de inteligência da Folha, que é o grupo de jornalismo de dados da Folha. Foi criado em julho de 2017.

#### **2. Breve histórico do JD no veículo. Iniciativa de quem?**

O jornalismo de dados não começou na Folha com o NIF (do núcleo de inteligência da Folha). A gente tinha aqui o Marcelo Soares, que foi um dos primeiros jornalistas do Brasil a trabalhar com dados. Por muito tempo ele trabalhou de uma forma meio individual. Depois o jornal criou um núcleo que era de jornalismo de dados, mas também de audiência, ou seja, o mesmo grupo cuidava dos relatórios de audiência do site. Acabou que audiência tomou a energia desse grupo e o jornalismo de dados ficou de lado. Então criaram o NIF, que é só mesmo para material jornalístico. A iniciativa da criação do NIF teve dois lados. Além do Marcelo Soares, havia outros jornalistas interessados na área, como o André Monteiro, que era do caderno Cotidiano e começou a programar muito bem. Então havia iniciativas individuais na redação. Mas o negócio ficou mesmo estruturado a partir de uma iniciativa da direção, da secretaria de redação que é a instância acima dos editores. O Vinícius Mota e o Roberto Dias, que são os secretários, nunca me contaram muito bem como foi, mas me parece que foi ideia do Vinícius e o Sérgio D'ávila, diretor executivo, e o Octávio Frias Filho, diretor de redação, toparam. Então, foram duas coisas paralelas. Mas até eu trabalhei com dados, mas foram coisas individuais. Importante ressaltar que, ao estruturar um núcleo, não saímos pegando quem estava na redação e sabia mexer com dados. Pegamos gente de fora e não são jornalistas, inclusive.

#### **3. Existe uma editoria ou sessão separada para dados/infográfico?**



Não. A gente já pensou algumas vezes, mas a direção desenhou assim, como se fôssemos um apoio para a redação, mas é mais que um apoio, pois apoio parece uma coisa meio help desk. De vez em quando até fazemos isso. Mas a ideia até esse momento é que não se distinga o material que é jornalismo de dados do que é jornalismo investigativo ou qualquer outra matéria. Então não colocamos um selo “jornalismo de dados”. Estamos fazendo jornalismo. Nesse caso estamos usando grandes bases de dados, mas no fim a ideia é que seja um material da Folha, quase indistinguível do resto do material. Quando o NIF participa, o cientista de dados assina junto com os repórteres, como qualquer jornalista assinaria.

#### **4. Qual o tamanho da equipe?**

Ainda bem que estamos tendo essa conversa esse mês, pois até mês passado só tínhamos um programador, aliás, um cientista de dados. Contratamos mais três cientista de dados. Então, a equipe sou eu e mais quatro programadores, ou cientista de dados, sei lá. Nem eles conseguem definir se eles são programadores ou cientista de dados. Mas a gente não faz tudo. Se é uma matéria de política, geralmente vem um repórter de política. De música, vem um de cultura. Eu coordeno o processo, mas a ideia é sermos uma equipe porosa.

#### **5. A equipe é dedicada somente ao JD?**

Sim. Eu e os cientistas de dados, sim.

#### **6. Fisicamente, a equipe se localiza próxima do resto da redação? Isso faz diferença no trabalho?**

Nós ficamos mesmo na redação, não tem sala separada. Tem perdas e ganhos. Numa sala separada poderia ter mais concentração, mas acho que é um ganho muito marginal. Estar na redação, as conversas acontecem mais e difundimos mais a cultura de dados. As pessoas sugerem mais pautas.

#### **7. Quem são os profissionais e quais as formações deles?**

Fábio Takahashi – coordenador do NIF, jornalista. Meus conhecimentos de programação são bem parcos, mas acho que são suficientes para conversar com os programadores, para fazer a ponte com o jornalismo, para fazer as perguntas certas para aquele banco de dados enorme. Eu tenho o mínimo de conhecimento para saber fazer essas perguntas.

Daniel Mariani – cientista de dados / programador, biólogo. Veio do Nexo. Foi o primeiro, e deu muito certo. Então, quando fomos crescer a equipe, procuramos dentro dos contatos

dele, por isso veio mais um biólogo. Eles trabalhavam com programação pesada na biologia. O Mariani procurava mutações genéticas que pudessem indicar esquizofrenia. O Guilherme estudou a evolução dos crânios dos mamíferos.

Tem um cientista social, da USP. E um outro graduado em análise de sistemas, mas fez mestrado e doutorado em Informática e Saúde, na Unifesp.

O Mariani já tem um faro jornalístico, pois ele veio do Nexo e já meio que aprendeu o processo. Mas é um processo, já que eles têm carreira fora do jornalismo. Os outros estão começando ainda, eles têm ideias interessantes, mas que precisam ser lapidadas até que vire um produto jornalístico.

#### **8. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

A regra é não ter muita regra. A gente tem ideias, dá uma pesquisada e vê se é viável, e apresentamos direto para a secretaria de redação, pois estamos imediatamente abaixo na hierarquia. Ou seja, estamos no mesmo nível das editorias, e não estamos subordinados a elas. Isso também depende, às vezes converso com as editorias. Em resumo, podemos ter a ideia, ou a ideia pode vir da redação. A partir daí, eu decido qual o cientista de dados pode ter mais afinidade com o assunto e ele faz a primeira exploração, para ver quanto tempo vai demorar, se é viável. Então, eu volto para o editor ou para a direção com um retorno. Sempre com uma conversa com a editoria, até o produto final, que geralmente é uma planilha. Então, nós tratamos os dados, estruturamos os dados, e as editorias é que escrevem as matérias. Eu dou umas opiniões. Sou a ponte entre as duas equipes. Mostro caminhos. Eles fazem a divulgação e a apuração, no caso entrevistar pessoas.

Nós fazemos visualizações poucas vezes, não é a regra. O que esperam da gente é que sejamos mesmo “da redação”. Esperam que façamos manchetes ou grandes especiais, e não infográficos legais e interativos. Isso vem de lambuja, mas se fizermos várias manchetes, a gente venceu a batalha. O interativo tem um problema de mão de obra. Então somos bem parcimoniosos em fazer interativos.

#### **9. Existe uma periodicidade definida para a publicação de “data stories”? Se sim, qual e por quê?**

Não. Tem coisa que fazemos em dois dias, tem coisa que fazemos em três meses, com todos os processos descritos acima.

### **10. Quais as ferramentas/softwarewares utilizados para produzir “data stories”?**

Basicamente R. Não foi nada planejado. Todos usam R, mas foi coincidência. Também SQL e até o Excel, às vezes. Com relação à visualização, é mesmo o departamento de arte que faz. Não sei quais são as ferramentas que eles usam.

### **11. Existe algum treinamento para a equipe?**

Está rolando algo bem legal. Quando começou o NIF, havia muita demanda da redação. Nem se tivéssemos 20 cientistas de dados, conseguiríamos dar conta. Então sugeri fazer um trainee de JD. É um curso de 4 meses, de dedicação integral, para jornalistas da casa e também jornalistas e profissionais de fora, por meio de um concurso, um para os internos e outro para os externos. Houve uma prova para os de fora, com noções de jornalismo, de programação (não necessariamente programação em si, mas noções de como funciona), habilidades de leitura de gráficos. Para os jornalistas da casa, a seleção foi mesmo essa parte prática. A gente esperava que haveria mais programadores interessados em aprender mais sobre jornalismo, mas só temos dois. Não é pago. Os da casa recebem como se fosse o dia de trabalho, ou seja, o salário deles, e o jornal está pagando substitutos para eles. Os de fora não pagam, mas não sei se recebem.

### **12. Do ponto de vista da recepção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

Eu acho que sim, pois a equipe cresceu 150%. Acho que não foi à toa. Mas não sabemos dizer se projetos mais interativos e com mais infográficos geram mais cliques.

### **13. Quais os desafios para a prática de jornalismo de dados atualmente?**

A Lei de Acesso a Informação não funciona tão bem. Tem dados que são públicos, mas eles escondem, ou colocam várias etapas para conseguirmos acessar. Tanto que contratamos um pessoal bom que conseguem “quebrar a caixa-preta”, mas não deveria ser assim. Foi até um dos motivos que contratamos gente que sabe de webscrapping. Só com um programa bom que conseguimos puxar esses dados. Acho que além disso, tem também um caminho para o JD se tornar mais relevante. É muito pessoal isso, mas acho que às vezes a gente se perde um pouco em mostrar como a gente sabe fazer, como a gente é bom em programação, e isso não é relevante para o público. Tipo, olha, fiz um programa incrível, tecnicamente fantástico, mas e daí? Temos recursos escassos em todos os lugares. Mas

isso é uma coisa que pessoalmente me incomoda, não sei se posso usar isso como uma regra para todo o jornalismo de dados.

#### **14. Qual a disponibilidade da empresa para investir mais em JD?**

Não posso nem cobrar mais crescimento nos próximos dois anos, pois crescemos 150%. Agora é amadurecer esse crescimento, pois primeiro temos que nos organizar e ver como as coisas vão funcionar.

## **II. ENTREVISTA COM LEONARDO CAZES – O GLOBO**

### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Leonardo Cazes, 29 anos, editor assistente do núcleo de dados e checagem. Entrei no Globo como estagiário em 2009, fiquei um ano e, em 2010, virei repórter. Primeiro na revista Magazine, um caderno jovem do jornal impresso, que acabou no início de 2012. Cobria mais a área de educação e, quando a revista acabou, continuei no site d'O Globo, na editoria de Educação, até setembro de 2012, quando recebi um convite para ir para o Prosa e Verso, um suplemento literário onde fiquei também até o caderno acabar, em 2015. Eu comecei a mexer com dados ainda quando era estagiário. Na época, O Globo tinha um programa de formação, uma oficina de dados, com o Fábio Vasconcellos, que hoje é professor e pesquisador. Ele era repórter da editoria de Cidades e dava curso de Excel, etc. No final do estágio, eu fiquei com uma pauta de dados sobre a segurança pública no Rio de Janeiro. Quando fui cobrir educação, me afastei um pouquinho, mas em 2011 voltei a trabalhar com dados da educação. Censo escolar, de educação superior. Foi meu primeiro contato com microdados. Houve até uma matéria que foi capa da revista Magazine, sobre presença de armas de fogo no ambiente escolar, pois essa é uma pergunta que os professores respondem no censo da educação básica. Só que eu não tinha conhecimento para levantar esses números. Na época eu pedi ajuda ao Pedro Markun, da Transparência Hacker, pois ele já tinha um script em Python. Eles rodaram os números para mim, e eu dei o crédito a eles na matéria. Mas nessa época eu já estava bem interessado em aprender, pois eu via que isso seria muito importante. Quando fui para o suplemento literário, me afastei dos dados. Quando o suplemento acabou em setembro de 2015, a equipe foi para

a editoria de cultura, e eu voltei a lidar mais com números, pois eventualmente eu era puxado para uma pauta de políticas públicas para cultura, mas de forma bem incipiente. Entre 2015 e 2017 fiz um mestrado em História, também não tinha nada a ver com dados e jornalismo. Em setembro de 2017, eu fui convidado para ser editor assistente do núcleo de dados e checagem, uma espécie coordenador de produção/chefe de reportagem dos dados e da checagem.

## **2. Breve histórico do JD no veículo. Iniciativa de quem?**

O núcleo de dados d'O Globo já passou por algumas fases. Quando entrei, em 2009, havia pessoas que faziam matérias com dados. Não existia um núcleo. Demorou muito tempo até que, acho que em 2012 ou em 2013, criou-se um núcleo de dados, ligado ao departamento de arte. Era o Fábio Vasconcellos, da redação, a Gabriela Alegro e o Daniel Lima, esses dois últimos editores de arte. Em março do ano passado houve uma grande reestruturação n'O Globo, com a integração com o jornal Extra, a mudança para um novo prédio, e agora com a revista Época está se integrando também. Nessa época, o núcleo de dados ganhou dois repórteres, o Marlen Couto e o Luís Guilherme Julião, ainda com o Fábio. Eles continuavam trabalhando junto com o departamento de arte, já havia o blog Na Base dos Dados, mas nessa reorganização eles assumiram a parte de checagem. E hoje, o departamento é de dados e checagem, e a prioridade do jornal é a checagem. Com a saída recente do Fábio Vasconcellos, em setembro, houve outra reorganização. O Fábio Gusmão, criador do blog "é boato, é verdade" do jornal Extra, criado em 2011 só com assuntos ligados à cidade do Rio de Janeiro, foi convidado para assumir o núcleo de dados e checagem. O Gusmão já era editor do Radar, que é uma espécie de super escuta que produz para todas as editorias do site, para dar as coisas em primeiro lugar. Então o núcleo de dados e checagem, mesmo tendo uma equipe separada, foi incorporado ao Radar. Nessa reorganização, com a entrada do Gusmão, eu fui convidado para ser editor assistente do núcleo de dados e checagem. Eu sou uma espécie de coordenador de produção/chefe de reportagem dos dados e da checagem. Acompanho as pautas, discuto.

A iniciativa de incorporar JD ao jornal veio muito mais dos repórteres do que da chefia. Teve até o apoio de um editor executivo, o Chico Amaral, que fez formação em Barcelona, era um entusiasta do JD. Mas quem correu atrás de capacitação, mesmo antes de ter um núcleo de dados, foram os próprios repórteres. Nunca houve um investimento pesado

institucional. Quando eu era estagiário, por exemplo, foi uma iniciativa da nossa turma de estágio. A gente sabia que o Fábio entendia do assunto e pedimos uma oficina com ele. Ele era repórter de cidades na época, e foi ele quem teve um papel decisivo no JD n’O Globo. Ele desenvolveu muitas pautas importantes antes da criação do núcleo, e depois houve a criação do núcleo e ele comandava.

**1. Existe uma editoria ou sessão separada para dados/infográfico?**

Não. Só existe o blog Na Base dos Dados. Fazemos matérias em parceria com outras editorias, e publicamos nessas editorias.

**2. Qual o tamanho da equipe?**

Cinco pessoas, todas jornalistas. Todos fizeram cursos por conta própria, como os da Abraji. Até a reestruturação de fevereiro, eram três pessoas. Sempre jornalistas.

**3. A equipe é dedicada somente ao JD?**

É dedicada à checagem e aos dados. Mas a prioridade atualmente tem sido a checagem.

**4. Fisicamente, a equipe se localiza próxima do resto da redação? Isso faz diferença no trabalho?**

Sim. Em setembro, o núcleo de dados e checagem, mesmo tendo uma equipe separada, foi incorporado ao Radar, que produz para todas as editorias do online. A ideia é “contaminar” o Radar com a cultura de dados e checagem. Antes, o núcleo de dados e checagem ficava perto do departamento de arte, Agora, perto do Radar, que é o noticiário mais quente do dia, e essa proximidade física também melhora bastante as pautas das matérias de dados, que ficam mais contextualizadas.

**5. Quem são os profissionais e quais as formações deles?**

Fábio Gusmão – editor Radar e núcleo dados e checagem

Leonardo Cazes – editor assistente núcleo dados e checagem

Marlem Couto – repórter

Luis Guilherme Julião – repórter (era trainee e virou repórter em setembro)

Gisele Barros – Trainee (contrato de um ano)

Estagiários (cada um fica um mês no núcleo, é rotativo)

Quando a equipe ficava perto da arte, tinha duas pessoas que estavam mais envolvidas: a Gabriela Alegro e o Daniel Lima. O Daniel continua trabalhando com a gente, mas ele é o editor assistente do departamento de arte de todo o jornal, então ele faz muita coisa.

Quando vamos fazer um projeto de dados maior, de fôlego, é com ele que trabalhamos, mas ele não é exclusivo para os dados.

**6. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

Podemos falar em dois fluxos diferentes. O primeiro quando somos demandados por uma editoria do jornal, por exemplo, Política. A história dos supersalários, por exemplo. Tínhamos uma notícia: o CNJ divulgou os salários de todo o judiciário. A editora de Política d’O Globo veio até mim e disse: vamos fazer algo sobre isso. A partir daí, eu e o Marlem começamos a baixar as planilhas. Tinha planilha bloqueada, tivemos que quebrar tranca de planilha... E então nós começamos a ver o que chamava atenção naquelas planilhas. Até sair a matéria, foi um processo que durou mais ou menos uma semana. Aí nós é que montamos algumas pautas em cima daqueles dados. Mostramos para a editora de Política para eles escolherem, entre os caminhos, qual o mais interessante. E fomos nós que fechamos a matéria até o fim. O Marlem escreveu, eu fiz uma primeira edição, acompanhei todo o processo de apuração, e entreguei para a editoria de política, e eles fizeram a edição final. Isso acontece com frequência. Outro exemplo: temos uma editoria chamada Sociedade, que inclui educação, ciência, saúde, tecnologia, história... Uma das editoras de Sociedade pediu ajuda para trabalhar com microdados, porque a Folha e o Estadão vinham fazendo uma série de matérias e estávamos ficando pra trás. Tivemos uma reunião, indiquei um caminho e trouxemos pautas sobre o assunto. Então, a editoria pediu ajuda e é uma coisa que, desde que fui para o núcleo, vi que isso acontecia pouco. Não somos prestadores de serviço das editorias, mas ao mesmo tempo temos que potencializar o nosso trabalho com as editorias, até porque temos só o blog.

O outro lado, que na verdade acontece mais, é: nós temos a ideia da pauta. Nós oferecemos para as editorias publicarem. Se elas quiserem, ótimo, senão colocamos só no blog. Todos os profissionais da equipe são encorajados a sugerir.

Trabalhamos com dados já estruturados. Raramente fazemos raspagem de dados. Fazemos mais como treinamento do que para usar nas matérias.

**7. Existe uma periodicidade definida para a publicação de “data stories”? Se sim, qual e por quê?**

Não. Como a prioridade é a checagem, os dados ficam em segundo plano. Quando surge algo grande de dados, aí fazemos. Agora que o Julião foi efetivado, neste ano, e temos a trainee, pode ser que façamos mais. E ainda tem um estagiário, mas como o estágio é rotativo, cada mês tem um diferente. Então depende muito do interesse do estagiário.

#### **8. Quais as ferramentas/softwarewares utilizados para produzir “data stories”?**

Atualmente, utilizamos algumas ferramentas para não precisar passar pela arte, já que a arte tem muita demanda de todo o jornal. Programas que usamos: Excel, 95% do tempo. Fizemos matéria recentemente usando o SPSS, da IBM, de estatística, para rodar dados maiores. Agora estamos tentando uma ponte com o setor de tecnologia do jornal para passar usar o SQL, para trabalhar com banco de dados. Na verdade, estamos no limite de linhas do Excel. Por exemplo, no caso dos microdados do censo escolar, que estamos fazendo com a editoria de Sociedade, só conseguimos rodar os microdados das escolas, pois tanto professores quanto alunos... tem muito mais de um milhão de alunos e professores, e o Excel só tem um milhão de linhas. A gente está começando a trabalhar com bancos de dados relacionais, e para isso o SQL é fundamental. Tivemos uma reunião com a equipe de tecnologia d'O Globo, que trabalha com análise de dados no viés de negócios. Há uma pessoa de Data Science, há um engenheiro de dados, e para eles nós não mexemos mesmo com dados, pois para eles a quantidade de dados que usamos é muito tímida. Então, o próximo salto do núcleo é capacitar a equipe para trabalhar com mais dados. Para produzir, usamos bastante o Tableau e o Datawrapper, porque são ferramentas que nos permitem “driblar” o pessoal da arte. Nada contra eles! É mesmo porque eles não têm tempo. É muito menos pela questão de o processo ficar fragmentado, o que realmente ocorre, pois eles finalizariam sozinhos. Mas eles têm demanda do jornal inteiro. Agora, bem recentemente mesmo, temos usado o Tableau para consumo interno. Inputamos uns dados no Tableau para facilitar a consulta para pensarmos pautas. Assim, as pessoas podem fazer os cruzamentos no Tableau, usando os filtros do Tableau, e todos têm acesso. No fim das contas, o que torna viável fazermos as coisas é o Tableau, pois se formos passar para a arte vai demorar no mínimo uma semana. Quando são projetos que consideramos especiais, aí a gente mobiliza a arte. Quando a gente quer uma visualização bacana... Mas para uma coisa do dia-a-dia, uma coisa mais rápida, aí é o Tableau. A gente tem investido em reportagens ligadas ao Hard-News, mais quentes. Então, se tiver



interatividade, vamos ter que mandar para a arte e esperar duas ou três semanas para ficar pronto. Nosso trabalho é estático por limitação de recursos mesmo.

#### **9. Existe algum treinamento para a equipe?**

Não sistematicamente. Esporadicamente há treinamentos. Hoje na equipe todos são jornalistas e todos já foram ao congresso da Abraji e fizeram oficinas de dados. Alguns também fizeram o curso do Knight Center. Todos têm interesse na área. Os próprios profissionais buscam suprir essas lacunas. No final de janeiro, eu, o Julião e o Marlen demos a primeira edição de um curso de checagem e dados para a redação d' O Globo. São treinamentos hands on. Foram dois dias, com um módulo de checagem e um de dados. O módulo de dados é focado no Excel, para as pessoas entenderem as possibilidades que o programa permite. Agora, haverá outro na semana que vem. O objetivo é difundir a cultura de dados e checagem na redação. A gente quer que eles entendam o que a gente faz e que possam fazer algumas coisas. As pessoas têm que ver que muitas coisas são simples. Não precisa escrever código para fazer algumas coisas. Até para que o núcleo possa focar no que realmente é mais complicado, mais complexo, que exija conhecimento de Python, SQL... Com o treinamento de 3 horas de Excel, os repórteres poderiam fazer muito do que fazemos atualmente.

#### **10. Do ponto de vista da recepção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

Eu acho que sim, pois a partir do momento que se cria um núcleo de dados, o jornal viu valor em alguma medida. Mas, por outro lado, eu acho que o jornal dava pouco espaço para pautas de dados no noticiário. O jornal achava importante, estruturou um núcleo de dados, mesmo que pequeno, mas deixou penetrar pouco no noticiário, e acabou ficando uma coisa apartada da redação. Isso a gente está começando a fazer agora. Um núcleo irradiador da cultura de dados para o resto da redação. Eu acho que havia um espaço para o jornal explorar mais e acabou não acontecendo. Mas é o que estamos tentando fazer agora, pois há um preconceito, um medo, com relação a dados, pois as pessoas pensam que é muito difícil. Então queremos mudar isso.

#### **11. Quais os desafios para a prática de jornalismo de dados atualmente?**

Com relação ao acesso aos dados, falando aqui do Rio de Janeiro... Quando se trata de governo federal, é bom e é fácil. O serviço da Lei de Acesso à Informação funciona no âmbito federal. Quando passamos para o governo do Estado do Rio de Janeiro e prefeitura do Rio de Janeiro, tudo é muito difícil. Em algumas áreas, já conhecemos alguns atalhos, mas no geral, via lei de acesso, demora cerca de seis meses. São duas realidades. É por isso que você vê muito mais matérias com dados federais do que regionais. E isso é no Rio de Janeiro, que é um estado que supostamente deveria estar mais estruturado para lidar com esse tipo de informação. Institucionalmente falando, é mesmo uma questão de investimento.

#### **12. Qual a disponibilidade da empresa para investir mais em JD?**

Neste momento, não. Nós já ganhamos recentemente uma pessoa a mais. Nosso objetivo neste momento é manter a equipe nesse tamanho.

### **III. ENTREVISTA COM DANIEL BRAMATTI – ESTADÃO DADOS**

#### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Daniel Bramatti, tenho 49 anos comecei a trabalhar com jornalismo em 91 ou 92. Eu trabalhava dentro de jornal, mas não era jornalista formado. Eu me formei, passei para a redação, trabalhei no jornal Zero Hora em Porto Alegre e em 94 fui trabalhar na Folha de São Paulo, sucursal de Brasília, onde fiquei até 2000, cobrindo basicamente congresso, alguma coisa de economia, fazendo esporadicamente um trabalho na redação como editor, ficava como coordenador das equipes e, nesse período, teve um intervalo em que eu participei de um programa que chamava correspondente bolsista, em que você saía do seu trabalho onde você estava alocado e ia para uma cidade do exterior durante 6 meses e recebia uma remuneração diferente. Eu fiquei em Buenos Aires. Em 2001, ainda pela Folha, eu fui para São Paulo e comecei a trabalhar na editoria de cotidiano. Era cobertura mais da cidade, dos assuntos locais, como saúde e educação. Em 2004, eu fui para a editoria de política como editor adjunto e fiquei até 2006, quando saí da Folha. Passei o ano de 2007 no Terra Magazine, que foi uma experiência 100% online. Era uma unidade que produzia

conteúdo para o Terra, mas não era exatamente gerida pelo Terra. Em 2008, eu fui para o Estadão e hoje [16 de março, data da entrevista] estou completando 10 anos de Estadão.

## **2. Breve histórico do JD no veículo. Iniciativa de quem?**

O Estadão Dados começou em 2012 por uma iniciativa do José Roberto de Toledo. Ele foi para um evento de jornalismo de dados do Knight Center para Jornalismo nas Américas, da Universidade do Texas em Austin, com os melhores jornalistas de dados do mundo. Foi o Toledo que introduziu o CAR no Brasil, ele foi pioneiro e, em 2011, o que ele viu lá na conferência, que era uma novidade nos Estados Unidos e que no Brasil não tinha, foi a integração entre o jornalista e os programadores. Foi aí que ele propôs ao Estadão investir numa unidade com essa característica. A aceitação foi rápida, pois informalmente ele já trabalhava nessa linha e dava muito certo, e então surgiu o Estadão Dados oficialmente, em meados de 2012, com o Basômetro, uma ferramenta interativa que permite medir o apoio dos parlamentares ao governo e acompanhar como eles se posicionaram nas votações legislativas. O Basômetro sempre teve essa característica de permitir que o usuário construa ou interprete a história do jeito que ele quer, sempre foi interativo. A ideia era que o sujeito pudesse analisar diferentes períodos, diferentes partidos, diferentes coalizões, e sempre foi uma coisa totalmente visual. A gente achou que ia ser uma coisa muito usada por políticos, mas não. A gente teve pouco feedback dos políticos. Quem usou muito foi o pessoal da academia. Um livro foi escrito por 10 cientistas políticos. Cada um fez um artigo baseado num recorte do Basômetro. Eles o chamavam de “parque de diversões de cientista político” (risos). Agora ele está há 6 meses sem atualizar, porque a Câmara dos deputados mudou a API de abastecimento e também porque algumas coisas eram manuais e agora já dá para fazer automatizado. Isso precisa de um programador e estamos sem programador. No início, nós tínhamos uns três ou quatro programadores que ajudaram a fazer a ferramenta. Recebemos reclamações de professores que usavam em sala de aula, e a gente fala: “espera mais um pouquinho que nós vamos resgatar”. A gente estava sem condições de contratar alguém para fazer esse trabalho, mas acho que agora vai sair.

Mas voltando para a história do Estadão Dados... O Toledo já me conhecia, porque quando eu trabalhava no Terra Magazine ele trabalhava numa outra empresa no mesmo prédio, então a gente se cruzava. E quando eu trabalhava na editoria de Estadão, eu, era editora-

assistente que é o segundo na hierarquia, mas teve um momento em que o repórter especial saiu do jornal e o editor fez uma proposta para uma pessoa, que recusou, e uma segunda pessoa que também recusou, e uma terceira, que também recusou. Então, acharam uma pessoa que tinha um perfil mais de edição, então me propuseram voltar para a reportagem. Mas fazia muito tempo que eu não fazia reportagem, eu já tinha perdido os contatos com as minhas fontes de Brasília, estava completamente desatualizado. Então eu investi em pegar dados de votações, de eleições, comecei a fazer coisas que pudessem render matéria sem que eu tivesse que correr atrás de fonte. E foi bem nessa época que o Toledo já estava pensando no Estadão Dados. Ele viu o resultado desse trabalho e me chamou.

Então, na primeira composição, além do Toledo e de mim, tinha também a Amanda Rossi que hoje está na BBC. Aliás, a Amanda tem uma trajetória muito legal de dados. Ela já fez trabalhos com dados na TV Globo, que é uma coisa difícil de fazer no jornalismo local em São Paulo. Tinha também o Diego Rabatone, que foi o primeiro programador exclusivo do Basômetro. Chegamos a ter uma pessoa com formação em Estatística, que era o Lucas de Abreu Maia, um jornalista com formação em Estatística. Aí saiu a Amanda, saiu o Lucas, e nós chamamos o Rodrigo Burgarelli e o Guilherme Duarte. O Rodrigo está hoje no Facebook e o Guilherme está no Jota [site de jornalismo jurídico]. O Rodrigo é um baita jornalista e tinha estudado até o terceiro ano de engenharia, então sabia bem de exatas e também aprendeu programação. O Guilherme tem uma formação em Direito, e depois estudou Matemática e sempre foi meio autodidata como programador. Ele não é jornalista, mas por causa do Rodrigo foi chamado para o Estadão e gostou tanto que ele está até hoje mexendo com isso. Então, naquele momento a gente tinha uma equipe muito boa: jornalistas com experiência em dados, programadores com noções de jornalismo... foi o nosso ápice.

Aí, no fim de 2016, o Toledo saiu. Foi quando eles formalizaram o Estadão Dados como uma editoria. Até então, o meu vínculo trabalhista era com a política, e o Burgarelli e o Guilherme também eram contratados por outra via. Neste momento, o jornal estava numa fase de contenção de despesas. A gente tinha o blog, mas prestava serviço para as editorias, principalmente para editoria de política até pela nossa origem. Nesse momento, o Estadão Dados quase acabou, pois eu fiquei sozinho durante muito tempo e eu não programo.

Depois eu conseguir chamar uma pessoa, a Cecília do Lago, que tem noções de programação, mas ela é jornalista, e não programadora. Como hoje a gente está sem essa capacidade de programação, então estamos suprimindo nossas necessidades pedindo profissionais “emprestados” da infografia. Essa matéria dos carros roubados, por exemplo, foi um trabalho de dados, mas capitaneado pela editoria de metrópole. A gente deu uma ajudinha de bastidores, mas eles fizeram quase tudo. Tem uma coisa interessante acontecendo, justamente pelo fato de não termos equipe: estamos fazendo muita parceria com a editoria de metrópole e de política, então hoje em dia o trabalho com dados está se disseminando na redação. As pessoas fazem o curso da Abraji, começam a mexer com Excel, e essa era uma coisa que a gente sempre estimulou. Hoje, a editoria de metrópole tem três jornalistas que claramente fazem jornalismo de dados. Na política, tem uma pessoa que está super interessada e também está se enveredando por esse caminho.

### **3. O fato de o Estadão ter sido pioneiro em JD no Brasil fez com que muitos veículos buscassem os profissionais que vocês tinham?**

A gente acabou perdendo alguns profissionais. Os primeiros programadores que trabalharam no Basômetro também trabalhavam para outras editorias e logo depois por coincidência eles acabaram saindo do jornal. Então tivemos que chamar o primeiro programador só do Estadão Dados, que foi o Diego Rabatone. Ele ficou bastante tempo com a gente, mas depois saiu. Agora, a parte visual depende muito de programação, as ferramentas que a gente colocou no atlas político do Estadão Dados envolvem algum tipo de programação, então está desatualizado, o Basômetro também e infelizmente o blog está meio que abandonado. A gente voltou a trabalhar para o papel, a gente produz muita reportagem hoje escrita e pouca coisa de visualização, por causa dessa falta de um profissional nessa área, por essa falta de recursos do jornal. E também tem outra coisa: a partir do momento em que o jornal disse que tem dinheiro para contratar, eu não sei se será tão fácil, porque encontrar alguém com perfil de um Rodrigo Burgarelli ou de um Guilherme, ou seja, um programador com interesse em jornalismo e que tenha uma formação de exatas. Esses dois nós perdemos, pois eles receberam propostas tão interessantes que eles nem colocaram hipótese de negociação, e eu acho que infelizmente o jornal também não poderia fazer uma contraproposta.

### **4. Existe uma editoria ou sessão separada para dados/infográfico?**

A gente não tem uma sessão no site. Houve uma sessão separada na época dos Panamá Papers, que foi um especial. Estadão Dados é uma editoria no sentido de que é hierarquicamente como se fosse uma editoria. Como eu e a Cecília não temos um lugar para publicar, o que produzimos é publicado nas outras editorias, mas é apenas jornalismo de dados.

#### **5. Então a equipe é dedicada somente ao JD?**

Sim. Só fazemos jornalismo de dados, mesmo publicando em outras editorias. Mas no momento, fazemos parceria com infografia, que também é o departamento de arte. Tem uma equipe mais focada para produção de infográficos para o papel e um grupo que produz mais para o online, e nesse grupo que produz para o online tem programadores.

#### **6. Qual o tamanho da equipe? Quem são os profissionais e quais as formações deles?**

A equipe já chegou a ter 5 pessoas. Por um tempo fui eu sozinho, e agora somos uma dupla: eu e a Cecília. Nós dois somos jornalistas com noções de programação.

Por causa do trabalho extra das eleições, devemos contratar um terceiro profissional, por volta de maio, que deve ser um programador.

Quando a equipe tinha 5 pessoas, éramos: o Toledo como coordenador, eu como repórter, o Burgarelli como repórter e programador, o Lucas de Abreu Maia como repórter com conhecimento de estatística e a Amanda Rossi, também repórter. O Guilherme veio depois. todos esses profissionais eram multifacetados, eles tinham competências que preencheram faziam a ponte entre programação e jornalismo.

#### **7. Fisicamente, a equipe se localiza próxima do resto da redação? Isso faz diferença no trabalho?**

A equipe fica na própria redação, mas numa sala envidraçada. A porta está sempre aberta e todo mundo entra a toda hora. Sem dúvida isso é bom.

#### **8. Como são definidas as pautas das "data stories"? Descreva as fases do fluxo de trabalho na produção de uma "data story".**

Normalmente, em 60 a 70% dos casos, as pautas partem de nós. Sugerimos e, a partir do momento em que as editorias aceitam, a gente produz e entrega. Trabalhamos com um projeto de curto prazo para entregar uma matéria por semana e paralelamente fazemos

alguma coisa que exija mais fôlego, que trabalhamos sem data de entrega. A gente tenta fazer pelo menos uma entrega por semana de uma reportagem que envolva dados com alguma produção de infográficos, mas são estáticos. Excepcionalmente, a demanda vem da editoria. O pedido do jornal é que nós ajudemos a editoria de política primeiro e, em segundo lugar, a de metrópole, porque são editorias muito propícias para trabalho com dados. Quando a matéria sai publicada na editoria ela só sai com o selo do Estadão Dados. Por exemplo, a editoria de política pediu matéria sobre o auxílio-moradia dos juízes. Brasília estava fazendo matérias sobre auxílio-moradia, mas era mais com a opinião de pessoas sobre o assunto. Quando saiu uma enorme base de dados com o salário de todos os juízes do Brasil, houve um monte de informações que só o Estadão deu. A base de dados era do Conselho Nacional de Justiça, estava disponível para todos, mas ninguém fez essa análise com foco nos dados. A Folha e o Globo fizeram matéria, mas somente somaram tudo. Nós fizemos uma análise mais detalhada dos dados.

**9. Existe uma periodicidade definida para a publicação de "data stories"? Se sim, qual e por quê?**

O jornal pede que a gente faça especiais de fim de semana, então geralmente as matérias são publicadas no domingo ou na segunda. Para publicar interativo não há periodicidade. Fazemos só quando o conjunto de dados pede interatividade.

**10. Quais as ferramentas/software utilizados para produzir "data stories"?**

A gente usa muito o Google Spreadsheet, pois mais de uma pessoa pode trabalhar ao mesmo tempo, além de estar hospedado na nuvem, então posso acessar de onde eu estiver. Ele tem quase todas as funcionalidades do Excel, pelo menos as mais importantes para a gente, mas ele tem uma limitação de 400000 células. Às vezes a gente tem mais do que isso, então a gente passa para o Libreoffice, que uma versão open source do Excel, que tem uma limitação de um milhão de linhas. Quando é algo ainda maior, a gente usa SQLite, que é uma versão simplificada do SQL, e eu e a cecília conseguimos fazer pois envolve só uma linguagem de programação instrumental para se comunicar com a base de dados. Para mapas usamos o Carto, que é a única ferramenta paga que usamos. Para a construção de gráficos simples, a gente usava o Infogram, mas agora usamos muito o Datawrapper, que é um software livre criado para jornalistas e mantido por um instituto que garante que vai ser sempre livre para jornalista. Inclusive ele está melhorando bastante nos últimos

tempos. Ele está com um investimento na área de mapas e talvez a gente possa até abrir mão do Carto. Quando a gente tinha programador, a gente produzia as nossas visualizações usando o D3.js. A infografia também usa o D3.

**11. Existe algum treinamento para a equipe (exemplo: coding para os jornalistas, jornalismo para programadores)?**

Não. Esses jornalistas da editoria de metrópole e política, por exemplo, aprenderam por interesse próprio. Fizeram cursos da Abraji, por exemplo. Eles percebem que sem esse domínio há coisas que eles não conseguem fazer.

**12. Do ponto de vista da recepção, qual o impacto das "data stories" para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

Eu não tenho nada de analytics disso, mas o que a gente percebe é que não é uma coisa que atrai massa ou volume, mas atrai um leitor muito qualificado. A gente acaba recebendo e-mail de parabéns de cientistas políticos, autores de livros, o que é um diferencial, que faz o sujeito fazer assinatura online. Todos os jornais fazem uma cobertura muito parecida e isso aparece como um diferencial do Estadão.

**13. Quais os desafios para a prática de jornalismo de dados atualmente?**

Muitas pessoas não estão familiarizadas nem mesmo com gráficos simples e às vezes as visualizações são elaboradas, exigem que você saiba como interpretar. Tem um déficit nesse sentido, é uma percepção minha. Também faltam nas redações equipes maiores, pois análise de dados é um trabalho de equipe. Se você não tiver três ou quatro pessoas com talentos e capacidades complementares, você não consegue avançar muito. Então precisa desse investimento de criar uma editoria com uma equipe e não com uma dupla. Outro desafio do jornalismo de dados é que não é uma coisa que você aprende uma vez e vai fazer sempre igual. Sempre exige atualização, mudança de base tecnológica. Outra coisa é a falta de profissionais. Os que existem são aqueles que se interessam por jornalismo, mas que geralmente são de outra área. As faculdades de jornalismo hoje têm pouco ou nada de jornalismo de dados. Os profissionais têm vindo de outras áreas.

**14. Qual a disponibilidade da empresa para investir mais em JD?**

Existe uma promessa de investimento que é para suprir essa necessidade que vai ter de fato no ano eleitoral, quando a demanda por dados é muito grande. Além disso, queremos



que em um mês no máximo o Basômetro esteja funcionando novamente. Isso será feito por um programador freelancer para cuidar só disso.

#### **IV. ENTREVISTA COM THIAGO REIS – G1**

##### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Meu nome é Thiago Reis, tenho 36 anos e já fui repórter, redator e editor. Comecei minha carreira na Folha em 2004. Trabalhei por 5 anos na Agência Folha. Desde então, já mexia com planilhas básicas de Excel e produzia reportagens a partir de bases consolidadas ou mesmo criando bases próprias. Fui para o G1 (editoria São Paulo) em 2009 – onde atuava esporadicamente com reportagens de dados. Em 2012, fui para a área de produção do portal, onde ajudei a implantar várias equipes do G1 pelos estados. Foi no final de 2013 que passei a trabalhar focado em dados com outra repórter, a Clara, que está na equipe até hoje. Era formado ali o embrião da editoria de dados do G1. Hoje, coordeno essa equipe.

##### **2. Breve histórico do JD no veículo. Iniciativa de quem?**

Sempre houve um incentivo para histórias guiadas por dados no G1 – especialmente durante o período eleitoral. Durante um tempo, os repórteres que trabalhavam com dados, no entanto, faziam isso de forma autônoma. Houve, então, treinamentos pontuais e a quantidade de projetos cresceu a ponto de surgir a necessidade de um grupo próprio, direcionado a isso na redação. A direção, então, percebendo a importância de uma área dedicada a dados, promoveu a criação de um núcleo, movimento consonante a outras iniciativas pelo mundo.

##### **3. Existe uma editoria ou sessão separada para dados/infográfico? Qual o tamanho da equipe? A equipe é dedicada somente ao JD? Fisicamente, a equipe se localiza próxima do resto da redação? Isso faz diferença no trabalho? Quem são os profissionais e quais as formações deles?**

Há um núcleo de dados no G1, que fica geograficamente ao lado da equipe de arte (que, por sua vez, conta com dois programadores). Além de mim, há quatro repórteres, todos

jornalistas, três em SP (sendo que um está praticamente focado em fact-checking) e um no RJ. Parte também tem conhecimento de programação.

**4. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

São definidas de várias formas. Em reuniões da equipe, em propostas individuais dos repórteres... Ela pode surgir a partir de um pedido via LAI, de uma análise em planilhas abertas ou mesmo de uma boa sacada. Muitas vezes pautas especiais de dados surgem do mero acompanhamento do noticiário, do chamado hardnews. Sobre o processo de produção, o primeiro passo é obter a base. Ela pode estar disponível em algum site (em formato de planilha ou de outro modo que necessite uma raspagem dos dados), pode ser requisitada por e-mail para a assessoria de imprensa de algum órgão ou mesmo solicitada via LAI. Com ela em mãos, dá-se início ao processo que talvez seja o mais importante de todos: a limpeza. Há ferramentas para isso. Mas às vezes é preciso colocar a mão na massa mesmo. Com a planilha pronta, é preciso fazer análises e entrevistá-la, como uma fonte qualquer. A partir das conclusões dessas análises, é que se começa a pensar na visualização e nas histórias que podem ser contadas a partir dela. Fontes são entrevistadas e personagens são procurados. No G1, é possível contar com a ajuda das afiliadas, espalhadas pelo país, para fazer cases verificados durante esse processo. Um exemplo: <https://g1.globo.com/economia/noticia/nordeste-em-emergencia-historias-de-uma-seca-sem-fim.ghtml>

**5. Existe uma periodicidade definida para a publicação de “data stories”? Se sim, qual e por quê?**

Não. Especialmente porque a decisão editorial hoje no G1 para a área é trabalhar em projetos de fôlego. Um exemplo é o Monitor da Violência (<http://g1.globo.com/monitor-da-violencia/>).

**6. Quais as ferramentas/softwarewares utilizados para produzir “data stories”?**

A gente trabalha muito com Excel, bloco de notas e as ferramentas do Google, como o Docs e o Spreadsheets. Mais que dados, as reportagens precisam conter histórias, personagens e outros elementos que capturem o leitor. A gente trabalha de diversas maneiras para melhor aproveitar uma base de dados: seja com vídeos ou animações, seja com

ferramentas de visualização como o Carto (para mapas) ou o Timeline JS (para linhas do tempo), seja com artes estáticas e recursos textuais que permitam uma leitura clara, direta e concisa. Um exemplo: <http://especiais.g1.globo.com/sao-paulo/2017/o-mapa-da-homofobia-em-sp/>

**7. Existe algum treinamento para a equipe (exemplo: coding para os jornalistas, jornalismo para programadores)?**

A maioria dos cursos é feita fora. Há diversos sites hoje (muitos deles grátis) que ensinam a programar, por exemplo. Internamente, há um esforço para que jornalistas de outras áreas tenham maior familiaridade com dados. Já houve intercâmbios com repórteres de outras editorias e um curso foi desenvolvido para as praças espalhadas pelo Brasil.

**8. Do ponto de vista da recepção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

Com certeza. Reportagens guiadas por dados dão credibilidade e agregam valor. Há inúmeros exemplos. O principal deles é o Monitor da Violência. Mas projetos de dados nem sempre se traduzem em audiências astronômicas. Há um desafio diário de aliar relevância a alcance. Existe sempre uma preocupação de que os infográficos interativos, por exemplo, sejam acessíveis e úteis. Em alguns projetos, especialmente os que tratam de questões que afetam diretamente a vida das pessoas, a resposta é maior. Um exemplo: <http://especiais.g1.globo.com/rio-de-janeiro/2018/fora-do-ponto/>. Os leitores também se interessam bastante por materiais de serviço/utilidade pública. Um dos principais projetos do G1, o Promessas dos políticos, nasceu com a checagem das promessas de apenas quatro prefeitos. No dia da publicação, vários leitores de diferentes partes do Brasil ‘cobraram’ projetos semelhantes em suas cidades. Hoje, além do presidente, há páginas dos 26 governadores e dos 27 prefeitos das capitais

(<http://especiais.g1.globo.com/politica/2015/as-promessas-dos-politicos/>)

**9. Quais os desafios para a prática de jornalismo de dados atualmente?**

Bases de dados abertas e registros confiáveis. A falta de transparência ainda é regra. Muitas vezes, mesmo pedidos via LAI não são cumpridos. Formatos não editáveis também são comuns. Poucos órgãos têm repositórios e bases de dados históricos disponíveis e/ou facilmente acessíveis, o que dificulta muito o trabalho.

## **10. Qual a disponibilidade da empresa para investir mais em JD?**

Até o início de 2017, eram apenas dois repórteres dedicados a dados no G1. A equipe dobrou de tamanho em um ano. Dentro da Globo, no geral, também há um investimento e uma integração maior entre os profissionais que mexem com dados. Cursos já foram feitos e reuniões sistemáticas entre equipes têm sido realizadas desde o final do ano passado.

## **V. ENTREVISTA COM RODOLFO ALMEIDA – NEXO**

### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Rodolfo Almeida, 25 anos, jornalista infografista do Jornal Nexo. Sou jornalista de formação, e durante a faculdade eu sempre me envolvi de alguma maneira com design gráfico, seja estudando por conta própria, seja fazendo projetos impressos, em quadrinhos... Durante a faculdade, fui estagiar no Estadão, durante 2 anos, e fiquei basicamente na TV Estadão, que é a editoria de vídeo que produz vídeo para toda a redação e também produz pautas próprias. Ali tive contato com a arte do jornal, que era uma editoria bastante próxima fisicamente, e lá tive muito contato com o modo de trabalho deles, os softwares usados. Assim que eu me formei, através de contatos em comum, surgiu a vaga no nexo. Meu cargo é de infografista, e em abril de 2017 eu faço dois anos no cargo. Lá que eu comecei a me envolver mais com dados, aplica a parte de design para dados. Tive que aprender um pouco de estatística e de outros conceitos de data science.

### **2. Breve histórico do JD no veículo. Iniciativa de quem?**

O Nexo surgiu em novembro de 2015. Não fui o primeiro infografista. Fui o segundo. Tinha um designer antes de mim, ele saiu e o Simon Ducroquet, que é o antigo editor de arte, me convidou para trabalhar lá. O Jornalismo de Dados sempre existiu no Nexo, faz parte do modelo de negócios e do modelo editorial do jornal.

### **3. Existe uma editoria ou sessão separada para dados/infográfico?**

Temos uma editoria chamada “gráfico”, uma sessão em que o leitor vai encontrar só os conteúdos que são explicações visuais de alguma coisa. Antigamente, o Nexo replicava uma estrutura de jornal impresso, com editorias de política, etc. Com o tempo, a diretoria e todo mundo percebeu que na verdade seria melhor mudar essa estrutura para combinar mais com o modelo digital. Então, em vez de editorias criamos núcleos. Tem o núcleo de gráficos e infografia, o de política e economia, o de especiais e interativos... Como infografista, o meu trabalho que antes era na editoria de arte, passou para uma área só de gráficos. Pode perceber que na sessão gráficos, só assinamos eu, que sou infografista, e o Gabriel Zanlorenssi, que é cientista de dados. E a equipe somos nos dois e agora temos um estagiário. Na parte de especiais e interativos, também há alguns trabalhos feitos pela área de dados.

**4. Qual o tamanho da equipe? Quem são os profissionais e quais as formações deles?**

Atualmente, somos três. Eu, que sou jornalista e infografista; o Gabriel Zanlorenssi, que é cientista de dados, graduado em Ciências Sociais pela USP (Universidade de São Paulo) e atualmente cursa mestrado em ciência política na mesma universidade. É pesquisador no Centro de Estudos de Política e Economia do Setor Público da Fundação Getúlio Vargas e no Núcleo de Estudos Comparados e Internacionais da USP; e o estagiário, que é o Gabriel Maia.

**5. A equipe é dedicada somente ao JD?**

É um núcleo de gráficos e pesquisa. Como o Nexo tem um modelo bem próximo da academia, tanto em termos de fontes e contatos, quanto da abordagem que é próxima do procedimento acadêmico, o jornal é estruturado para ter esse núcleo formado por jornalistas e pessoas que têm uma carreira acadêmica, como o Daniel Mariani, que é biólogo, e que agora está na Folha. Hoje, como somos eu, o Gabriel Zanlorenssi, que é cientista, e o estagiário, que é o Gabriel Maia, nossa produção principal é de gráficos, mas também colaboramos com a pesquisa para o restante da redação e tocamos projetos especiais e/ou interativos quando eles envolvem dados.

**6. Fisicamente, a equipe se localiza próxima do resto da redação? Isso faz diferença no trabalho?**

Sim, fica todo mundo junto. Somos só 32 pessoas, contanto administrativo, manutenção. É uma equipe pequena, ficamos todos numa mesma sala.

## **7. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

Tem algumas pautas que a gente começa porque já temos uma base de dados. Por exemplo, dados do IBGE, do Banco Mundial, ou então saiu um relatório. E tem outras pautas, em que a gente tem a curiosidade sobre um assunto e então vamos procurar dados a respeito. E às vezes a gente acha coisas que não era o que queríamos, e aí não vai para frente. Nós fazemos raspagem de dados, mas como último recurso mesmo. O ideal para trabalhar é ter os dados estruturados ou semiestruturados. Todos os processos de dados, a gente faz. Fazemos através do R. Algumas secretarias de governos têm equipes que facilitam bastante, e já achamos tabelas prontas. Outros órgãos já dificultam, e aí temos que raspar. É caso a caso. Por exemplo: a gente fez um gráfico sobre número de alunos nas salas de aula. Pegamos dados do Inep, via Ministério da Educação. Nesse caso, foi bem fácil de baixar os dados. Olhamos o que tinha de interessante. Eles separam tipos de escola – privada, municipal, estadual, federal – e também por ano de ensino. E eles têm média de alunos por sala. Aí eu e o Gabriel juntos olhamos essa tabela, e pensamos a que conclusões se pode chegar, e já começamos a pensar quais as visualizações seriam boas. Podemos descobrir a média de alunos por sala para cada tipo de escola. Com essas dúvidas, começamos a esboçar as primeiras versões dos gráficos, bem rudimentares, só para termos uma ideia. Então separamos quais serão interessantes. A partir daí, começamos a montar uma história e dar uma cara. No geral, a primeira coisa que pensamos é qual vai ser o título, ou seja, qual o guarda-chuva que vai empacotar isso. O leitor vai clicar nisso esperando ver qual o número de alunos nas salas de aula. Uma parte importante do nosso trabalho é educar o leitor a aprender a ler gráficos. Ainda tem muita gente que não domina esse conhecimento visual. Nessa reportagem, usamos um “truque” para criar uma animação, que é um gif. Os anos vão mudando automaticamente.

Uma coisa que eu vejo que é bastante importante é a combinação, em uma mesma pessoa, ou em duas pessoas na mesma equipe, de jornalista e designer. É fundamental, porque essa coisa de jornalista escrever e designer desenhar, de forma separada, não é a melhor opção. Isso foi uma coisa que eu tirei muito do Simon Ducroquet. Ele ajudou a montar o projeto gráfico do Nexo, e também é jornalista e designer de formação. Então ele tem a cabeça bastante voltada para essas duas coisas. Agora ele voltou para a folha, mas só para

infografias para especiais. Ele que me influenciou bastante nessa coisa de já ir escrevendo enquanto se está desenhando e ir desenhando enquanto se está escrevendo, e a coisa fica muito mais fluida.

Então todas as fases da reportagem de dados são feitas por nós três. Eu e os dois “Gabrieis”. Vamos partir sempre de uma tabela, ou de uma API para tirar dados do Twitter por exemplo, e aí vamos conversar sobre a visualização, e então vamos ou criar uma visualização direto no R, ou então eu mesmo vou pegar essa tabela e criar algo no Illustrator ou com ferramentas de geoprocessamento como QGIS.

Normalmente, ao começar a olhar os dados, se eles forem muito recheados, podemos entender que valem uma especial. Se forem legais de o usuário explorar por conta própria, vamos entender que valem uma interativa. Se quisermos só mostrar uma coisa, vamos tentar encaixar como um gráfico. Esse feeling de qual será o uso que o leitor vai fazer daquele dado é bastante do jornalista.

#### **8. Existe uma periodicidade definida para a publicação de “data stories”? Se sim, qual e por quê?**

Publicamos na sessão “gráficos” todas as segundas, quartas e sextas. Antigamente eram dois por semana, e os gráficos eram mais longos, com mais texto, mas dava muito trabalho e pouco retorno. Então, por decisão editorial, decidimos ter gráficos mais curtos e direto ao ponto, e diminuindo o tempo de produção conseguimos chegar ao cronograma de 3 gráficos por semana.

Os interativos, a gente faz junto com a equipe de tecnologia, que são desenvolvedores e que trabalham com Java Script. Não tem periodicidade. No geral, tentamos ter um grande interativo por mês ou a cada dois meses. Uma pauta dessa pode durar de 3 semanas a um mês e pouco, pois precisamos de mais gente, como a equipe de tecnologia.

#### **9. Quais as ferramentas/software utilizados para produzir “data stories”?**

As que a gente mais usa são R e Illustrator, para criar visualização, anotações, e deixar o produto mais pronto para publicação. Illustrator sobretudo para parte mais visual e o R para os dados. Quando eu preciso fazer um mapa com dados ou um mapa mais complexo, uso o QGIS, que é um programa de geoprocessamento, é open source, é fácil de usar e não precisa saber programação. É recomendável saber um pouco de Excel para trabalhar com

as tabelas que serão usadas no QGIS. Se usar integrado com o Illustrator, dá para ter um resultado bem sofisticado visualmente. Tem também algumas ferramentas menores online que às vezes usamos, como o RAWgraphs. O Tableau não temos muito interesse em usar, pois apresenta soluções mais de gaveta, prontas, além de não conversar muito bem com o site. O D3.js nós usamos apenas para fazer visualizações interativas, pois a curva de aprendizado é um pouco maior e exige mais *fine-tuning*, de mexer em código. Nas pautas do dia a dia a gente não usa muito.

Na sessão de interativos, tem por exemplo esse: descubra qual cidade brasileira tem o clima ideal para você. Foi feito todo com D3.js.

#### **10. Existe algum treinamento para a equipe?**

Treinamento específico não tem, justamente porque temos uma equipe bem pequena. Resolvemos essas questões no convívio. Eu, trabalhando no Nexo, aprendi um pouco de Python, um domínio mesmo instrumental só para mexer em alguma tabela que eu precise.

#### **11. Do ponto de vista da recepção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

A gente mede o feedback dos leitores, sobretudo com redes sociais. O público do Nexo é bastante especializado, então sempre temos um feedback. Mas a gente percebe sobretudo por acessos. Os gráficos mais curtos, mais diretos ao ponto dão mais resultado. Tem gente que acompanha o Nexo só por causa dos gráficos. Então às vezes vale muito mais abordar um assunto através de gráficos do que de textos, pois percebemos que o leitor vai ter uma experiência melhor. A gente vê sempre que tem gráficos entre as matérias mais acessadas.

#### **12. Quais os desafios para a prática de jornalismo de dados atualmente?**

Ainda há muito a avançar na questão da disponibilização de dados no Brasil. Essa semana mesmo estávamos trabalhando com dados do IBGE que é o maior órgão de estatística do país, e mesmo assim temos muita dificuldade, por exemplo, de achar qual foi a metodologia de como foi feita aquela tabela, o que aquele dado está mostrando. Então muitas vezes as coisas são pouco transparentes. Isso quando o problema não é outro de o dado simplesmente não estar disponível, ou estar disponível em baixa qualidade, com pouco rigor metodológico. Às vezes falta o rigor de qual a melhor maneira de divulgar um dado, de forma que o público vá entender. A coisa positiva é que cada vez mais está mais fácil



para as pessoas aprenderem a trabalhar com dados, por exemplo com esses cursos online de data science. Tem muita gente fazendo esse trabalho por conta própria, sem ser para um veículo de comunicação. Sobre a Lei de Acesso a Informação, nós não usamos muito, pois não tivemos uma necessidade imediata, mas eu vejo por parte dos jornalistas que é comum o pedido não ser retornado. Mas também há erros por parte dos solicitantes, que muitas vezes pedem para os departamentos errados. A gente fez um gráfico só sobre lei de acesso e vimos que muita gente faz o pedido errado.

### **13. Qual a disponibilidade da empresa para investir mais em JD?**

Não sei dizer, mas a parte de Jornalismo de Dados é um pilar fundamental no Nexo. Acho que existe sim a possibilidade de se expandir a equipe, pela importância do setor, mas sinceramente não sei se vai acontecer.

## **VI. ENTREVISTA COM SÉRGIO SPAGNUOLO – VOLT DATA LAB**

### **1. Nome, idade, breve descrição da carreira e como chegou ao jornalismo de dados.**

Sérgio Spagnuolo, 33 anos, editor-chefe e fundador Volt Data Lab. Eu comecei minha carreira jornalística trabalhando com jornalismo internacional e econômico, principalmente econômico. A gente trabalhava muito com dados, mas não no nível que eu trabalho hoje. Em 2013, eu saí das redações e fui trabalhar num projeto da ONU no Rio de Janeiro, como coordenador de comunicação. Foi no mesmo mês em que entrei lá que estourou a coisa do Edward Snowden, do vazamento da NSA, e de repente o jornalismo de dados, o jornalismo mais computacional digamos assim, começou a ganhar notoriedade. E no Brasil isso era incipiente, e eu comecei a pegar gosto por isso, comecei a estudar, a ir atrás de aprender programação. Eu não sabia nada de programação naquela época. Eu sempre usei muitos tutoriais, até mais do que cursos, pois cursos pressupõem que tem alguém te ensinando. Tem tudo explicado na internet. É muita tentativa e erro, precisa de muita paciência. Eu vejo muitas pessoas que já começam melhor que eu, mas desistem. Mas tem que insistir, pois você não acerta de primeira. Mas eu não sou programador, eu sou jornalista. Eu sei programar por consequência desse meu interesse, mas não é minha principal área. Depois voltei para esse projeto da ONU e, após um ano lá, eu quis voltar a trabalhar com jornalismo e aplicar o conhecimento que eu estava adquirindo. Aí, em 2014,

eu estava somente fazendo trabalhos freelance (durante a copa do mundo, fui repórter de eleições do Yahoo...) e depois que acabou essa etapa cheia de notícias por conta dos grandes eventos fiquei meio sem saber o que fazer. Se eu voltasse para redação seria mais difícil explorar o que eu estava fazendo, pois tinha pouca abertura para isso. Eu trabalhei muitos anos na Reuters, adorava o lugar e todo mundo, mas lá você tem que cuidar do noticiário do dia a dia, e o que eu faço mais é análise, raspagem de dados, tem a ver com programação, arquitetura da informação, montar banco de dados... Se eu fosse para as redações, provavelmente eu não faria isso e era uma coisa que eu queria explorar. Daí, continuei fazendo freelances e montei um pequeno blog, chamado Volt Data Lab, onde eu publicava minhas aventuras nessa área.

## **2. Como e quando surgiu o Volt Data Lab?**

Como eu disse, no fim de 2014, criei um blog, o VOLT DATA LAB e comecei a publicar as coisas que eu estava fazendo. A partir de um dos projetos, a Conta dos Passaralhos, que mostra as demissões de jornalistas no Brasil desde 2012, surgiram pedidos de parcerias e serviços. Nasceu daí a ideia de transformar o Volt em uma agência de jornalismo, produzindo conteúdo a partir de dados. É uma agência de jornalismo independente focada em investigação, análise e visualização de dados, desenvolvendo projetos em diversas áreas, como política, economia, direitos humanos, mídia e tecnologia.

## **3. Como você conceitua jornalismo de dados?**

Para mim, jornalismo de dados, que é uma abreviação de jornalismo orientado por dados, é uma prática dentro do jornalismo em que a sua matéria é guiada principalmente por estatísticas, não necessariamente pela prosa ou pelas fontes, mas sim pela análise quantitativa. Eu gosto muito da definição do Nate Silver, que fala que o jornalismo de dados é a aplicação do método científico no noticiário. Talvez o que eles fazem no FiveThirtyEight esteja mais ligado a estatística do que ao jornalismo propriamente, mas eu gosto dessa definição.

## **4. E dentro desse conceito, que o papel têm a visualização e a interatividade?**

A visualização de dados é muito importante pois quando você está tratando um grande volume de dados você tem que dar um sentido para aquilo. A visualização é a melhor forma

de mostrar conhecimento para as pessoas. Sempre que a gente pode e tem orçamento eu contrato um designer.

### **5. Você acha que o bom jornalismo de dados tem visualização e interatividade?**

Acho que o bom jornalismo de dados é o que traz matéria boa. A visualização é o resultado disso. Se é possível chegar a conclusões boas sem representar visualmente, tudo bem. Confundem o jornalismo de dados com fazer gráficos. Jornalismo de dados é produzir reportagem a partir de dados e chegar a matérias exclusivas ou análises relevantes. Os gráficos são um produto desse processo. É um jeito de representar a informação e mostrar para o leitor de um jeito melhor do que em tabelas e textos.

Quanto mais se agregam dados e categorias, menos você precisa de representação gráfica, pois você está diluindo informações em poucas coisas. Exemplo: uma planilha com um milhão de salários de servidores federais. Olhando para aquilo você não vê sentido. Mas se você pega salários por ministérios, ou por área, ou por pessoas com mestrado, etc, você dilui em poucas categorias e fica mais fácil explicar em forma de texto. É o que a Folha tem feito muito. Mas a visualização é mais impactante, tem um papel importante para ajudar o leitor a explorar os dados. Aí é que entra a parte do Interativo. Quando a visualização é interativa, ela é a própria narrativa, a pessoa tem que saber a ler o gráfico, exige um esforço do leitor também de querer explorar.

Na visualização que o Estadão Dados fez com os furtos de carros em São Paulo, por exemplo, eu achei um veículo meu que foi furtado em 2008. Então, para mim, aquilo teve um significado. Eu explorei a base de dados, como leitor, e me achei ali. É significativo você permitir que as pessoas façam isso. A interatividade tem esse papel. Tem muita gente que chama de interatividade você passar o mouse em cima e aparecer um número que já está no gráfico. Isso não é interatividade. Isso é útil, às vezes, para ficar mais bonito ou mais fácil de ver, mas não é interatividade, é um recurso para não poluir a visualização, é uma animação. Para mim, interatividade é quando a pessoa consegue explorar a base de dados a partir do visual que você criou. Mas não é obrigatório. O Aos Fatos, por exemplo, que eu colaboro, faz um trabalho muito bom a partir de dados e quase não tem interatividade e muitas vezes não tem gráfico. Então acho que dá para fazer excelente jornalismo de dados sem gráfico, mas o gráfico ajuda. Chama a atenção, é legal, sei lá, eu gosto, eu sou nerd (Risos).

## **6. Quais as ferramentas/softwarewares que o Volt utiliza?**

A base do Volt, hoje em dia, para análise de dados é o R. Já fizemos projetos em Jungle, que é um framework de Python, já fizemos em PHP. Usamos muito o SQL. Construir um banco de dados é outra coisa, mas em questão de análise de dados a gente usa R, principalmente. A Renata Hirota, minha analista, gosta de Python, que é o que ela estuda na graduação de estatística, que ela está fazendo, mas usamos muito o R. Para visualização, usamos muito o D3 para fazer a parte interativa. Mas também fazemos muita coisa estática, em R mesmo, eu gosto e vejo muito valor. Interativo tem que ser aquele conjunto de dados que pede para ser interativo. Como eu disse, eu não gosto de colocar botãozinho que quando você passa o mouse e aparece porque é um trabalho grande para pouco resultado. Mas quando você tem um interativo legal, como o que eu fiz sobre a rejeição de vistos brasileiros nos EUA, para a Piauí, aí é diferente, pois o conjunto de dados pedia a exploração.

## **7. Qual é a equipe do Volt? No site, diz que é uma *Lean startup*...**

Sim, como o Volt atua num tipo novo de segmento em comunicação no Brasil, adotamos o conceito de lean startup. Então, trabalhamos com uma rede de colaboradores, freelancers, o que nos permite acionar os profissionais necessários para cada trabalho, de acordo com a demanda. Hoje em dia, eu fico na coordenação de projetos e na parte metodológica, e eu passo a parte bruta da programação para outras pessoas, freelancers. O meu mestrado é em relações internacionais, e a fellowship que fiz nos EUA é de empreendedorismo e me ajudou muito nesse projeto.

Quando começou, era só eu. Na verdade, o Volt existe desde 2014, mas só começou a decolar mesmo e ser minha principal fonte de renda e ter minha dedicação quase exclusiva a partir do ano passado (2017). Antes disso, eu fazia freelances, pois precisava pagar as contas. A partir do ano passado, surgiram oportunidades, comecei a ser comercialmente mais ativo também, indo atrás de oportunidades. Mas sempre fui basicamente sozinho. Aí no final do ano passado, em setembro, fiz um processo seletivo de bolsas, e vamos fazer de novo esse ano. Foram 360 candidatos. Foi uma loucura, eu não esperava isso tudo. Esperava 20 ou 30. Na verdade, eu precisava de alguém para me ajudar e ao mesmo tempo eu queria desenvolver a comunidade, pois precisamos de mais profissionais no Brasil trabalhando com jornalismo de dados. E agora felizmente isso está acontecendo, inclusive

com uma diversidade, equilíbrio entre homens e mulheres, mais pessoas de minorias começando a atuar nessa área.

O que aconteceu no caso da bolsa foi que, no processo seletivo, a Renata Hirota, que é jornalista e está fazendo graduação em estatística, se destacou muito. Muitas pessoas se destacaram, na verdade. Mas ela foi a melhor. A primeira etapa do processo foi currículo e depois foi uma prova bem básica, para ver se as pessoas sabiam mexer com tabelas. Nisso, selecionei 10 pessoas para entrevistar. E ela passou. E depois que acabou a bolsa de 3 meses (eu pagava uma quantia e também um curso online de programação) eu a chamei para continuar comigo, e agora ela é minha colaboradora fixa.

Também tem o Felipe, que é o coordenador de tecnologia do Volt. Ele trabalha na parte de modelagem de banco de dados e back end. Ele é autodidata, não tem formação acadêmica.

#### **8. Você conhece alguma outra agência de jornalismo de dados no mundo?**

Nesse modelo do Volt, eu não conheço nada igual. Existem muitos empreendimentos relacionados a visualização de dados ou banco de dados, mas uma agência de jornalismo de dados eu nunca vi. Há alguns lugares, como o open data city na Alemanha, mas não fazem jornalismo de dados como negócio. Pode até ser que exista, mas eu não conheço.

#### **9. E qual a vantagem de uma agência de jornalismo de dados. Como você viu nisso uma oportunidade de negócio?**

É uma área que já se provou. A gente consegue muita informação relevante, muito furo de reportagem, muita análise legal, visualizações bacanas, mas ao mesmo tempo é muito caro se manter uma equipe de jornalismo de dados, com programador, designer... se você vai desenvolver projetos esporádicos, você não vai manter 5 pessoas. Então, muitas empresas de jornalismo veem no Volt uma forma de desenvolver projetos que não teriam condição de fazer antes. A gente está fazendo um projeto agora com um veículo de comunicação que não posso revelar, e eles têm uma pilha de dados para trabalhar e não têm equipe. Então nos contrataram. Se eles pagassem cinco pessoas sairia mais caro do que pagar para nós fazermos.

#### **10. Como são definidas as pautas e parcerias? Qual é o fluxo de trabalho do Volt?**

Geralmente, os clientes vêm até nós para desenvolver projetos, porque viram projetos que fizemos por conta própria. Ou então fazemos um projeto e depois tentamos vender. A

matéria sobre a utilização de redes do governo para alterações na Wikipédia foi feita por conta própria e fizemos uma parceira com a Piauí depois para divulgar. E estamos para publicar uma com o The Intercept que foi consequência disso também. Muitos projetos próprios nossos tem a ver com a própria comunidade de jornalismo de dados. Lançamos recentemente um site de curadoria, temos a conta dos Passaralhos, enfim, temos projetos próprios que servem para alimentar essa interlocução com a comunidade de jornalismo de dados, que está crescendo.

#### **11. O grupo do whatsapp #ddjbr foi uma iniciativa sua?**

Sim, eu sou o segundo mais ativo (risos). O Guilherme Jardim, cientista de dados do Jota, que passou pelo Estadão Dados, é o mais ativo. Tinha 12 pessoas quando o criei, no ano passado. Agora tem cerca de 130. Tem gente de todos os lugares, veículos, e existe uma solidariedade, uma confiança, uma colaboração. A comunidade se ajuda. Se você tem uma dúvida, joga lá e o pessoal ajuda. É uma área que está dando resultado. Até a Piauí, que era uma revista de prosa, agora publica gráficos. E as empresas estão começando a investir. E se elas não querem investir em equipe própria, porque é caro, eu estou aqui para isso! (risos)

#### **12. O Volt dá cursos?**

Sim, já demos mais de 10 cursos. São cursos presenciais. Ensinamos a trabalhar com planilhas em nível avançado. Ensinamos a trabalhar grandes bases de dados, usando SQL. E fundamentos básicos de jornalismo de dados, estatística. A academia não está dando conta de suprir essa necessidade.

#### **13. Quais os desafios para o jornalismo de dados no Brasil?**

Acho que principal dificuldade hoje no Brasil é a abertura para conseguir produzir conteúdo assim. Sobre o acesso a informação, as pessoas reclamam muito, mas acho que está melhorando, é assim mesmo. E outra: sou muito favorável às pessoas produzirem os próprios dados, e não dependerem só do governo ou empresas para mandarem as informações. É o que fazemos no Atlas da Notícia. Fizemos crowdsourcing, pegamos dados do governo também, de associações, e misturamos tudo. Então a gente está produzindo uma nova série estatística a partir de dados que nós mesmo estamos coletando. A conta

dos passaralhos é assim também. Foi o que a Folha fez com a matéria da música no YouTube. É o tipo de coisa que mais gosto.

## **VII. ENTREVISTA COM SOFIA MIGUEL ROSA E RAQUEL ALBUQUERQUE – EXPRESSO**

### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

**Sofia Miguel Rosa**, infografista. Eu comecei a trabalhar no Expresso em 1999 quando eu ainda estava a tirar o curso de Design. Eu percebi que gostava de design editorial, não especificamente infografia. Ainda estava a estudar, mas enviei o currículo para uma série de jornais e revistas e recebi um telefonema do Expresso para uma entrevista para o departamento de informática. E eu confesso que não sabia o que era infografia, mas disse “claro que sim”. Na entrevista, me pediram para mostrar alguma coisa sobre algum trabalho sobre gráficos e a única coisa que eu tinha eram gráficos de Excel feitos nos trabalhos escolares, nada que fosse digno de ser mostrado. Eles pediram para fazer um exercício, eu fiz, eles gostaram e fiquei. Terminei o curso, e a ideia era depois de ir para o exterior, mas acabei por ficar e estou aqui desde 99, portanto há 19 anos.

**Raquel Albuquerque**, repórter. Eu estou no Expresso há 4 anos. Estudei Ciências da Comunicação, com vertente em jornalismo. Fui fazendo alguns estágios e em 2012 surgiu a possibilidade de uma bolsa para um projecto de investigação de jornalismo computacional que envolviam o Instituto Superior Técnico, a Universidade do Porto e a Universidade do Texas em Austin, além do jornal Público. A ideia era trazer o conceito do jornalismo computacional para Portugal e montar uma redação de jornalismo computacional no Público. O projeto, chamado REACTION, foi financiado pela FCT e envolvia investigadores de doutoramento de diversas engenharias principalmente engenharia informática e um jornalista, que era eu. O António Granado na altura estava no Público e ele montou esse projeto com um professor de engenharia informática do Técnico. Portanto, eu tive essa bolsa por 2 anos, e aplicava a computação às práticas jornalísticas, ou seja, usava ferramentas informáticas para interpretar e resumir informação, por exemplo. Ou então criava modelos para apresentar conteúdos dinâmicos e interactivos. Era testar as ferramentas e analisar o impacto prático no processo de redacção. Eu não era paga pelo Público e sim pela FCT, mas a determinada altura eu estava

trabalhando na redação do Público com uma grande vantagem que era não estar presa na rotina diária do jornal. Eu fazia exclusivamente trabalhos com a parte mais computacional, que envolve por exemplo a análise de redes sociais, é uma coisa muito mais abrangente do que propriamente estar a trabalhar com dados. Mas depois pude também testar efetivamente o jornalismo com dado, e foi quando eu tive alguma formação em jornalismo de dados através de cursos online, uma coisa mais autodidata. Então, então durante aqueles dois anos fiz tudo o que havia que podia ser feito no âmbito deste projecto de investigação. Quando a bolsa estava a acabar, houve a possibilidade de ficar no Público, mas ao mesmo tempo surgiu uma proposta aqui no Expresso e eu optei por vir para cá. Nos últimos 4 anos, em umas alturas mais em outras menos, tenho vindo a tentar fazer coisas com jornalismo de dados, e não a parte computacional apenas.

## **2. E como você diferencia esses dois trabalhos, que seriam o jornalismo computacional e o jornalismo de dados?**

**Raquel** - Não há fronteiras muito definidas. O que eu sinto é que o jornalismo computacional na forma como eu fazia era uma coisa muito mais computação do que o trabalho do jornalismo de dados que eu faço atualmente sem conhecimento de programação. Hoje faço muito mais análise de dados, uso ferramentas mais simples como o Excel, ferramentas de limpeza de dados, como o Google Refine. O que eu sentia no projecto era que o ponto de partida eram coisas muito mais técnicas e às vezes era muito difícil escrever uma notícia com base naquilo. Computacionalmente era uma coisa extraordinária, mas jornalisticamente eu acho que aquilo tinha pouco uso. Mas a visualização era parte essencial daquilo.

## **3. Breve histórico do JD no veículo. Iniciativa de quem?**

**Sofia** - Eu acho que começou por volta de 2003 ou 2004, e também com as nossas idas Malofiej em Pamplona. Foi quando começou a haver uma mudança no sentimento e na própria identificação da profissão. Até aquela altura nós não nos considerávamos jornalistas, éramos designers. A partir daí começou-se a assumir que a infografia tinha uma componente jornalística e que deveria seguir os princípios básicos, éticos e deontológicos do jornalismo, e por e por aí afora. Eu acho que foi essa a grande mudança, quando começamos a nos ver mais como autores e menos como executores ou decoradores. A parte plástica continua obviamente a ter um papel forte, mas a partir daí começamos a nos



considerarmos também autores efetivos da informação jornalística. É mais do que um cargo, é uma questão de como olhar o trabalho.

**Raquel** - Há 4 anos, o que motivou a minha vinda para cá foi o jornalismo de dados, mas o que acontece há anos no jornalismo é o uso de dados para fazer jornalismo comum, ou seja, é trabalhar com estatística, é um jornalismo *com* dados. Isso porque o jornalismo de dados é uma coisa difícil, pois leva tempo, os dados não estão disponíveis, etc. É quase impossível fazer jornalismo de dados no seu conceito original na velocidade de uma redação. Então isso nos força a fazer um trabalho de médio ou longo prazo meio de bastidores, por iniciativa nossa, e às vezes é muito difícil de conciliar com todo o resto do trabalho diário. Nos trabalhos que eu vou fazendo, e reconhecendo que muitas vezes eu não estou a fazer jornalismo de dados e sim esse jornalismo com dados, eu tento pelo menos que aquilo que eu estou a fazer, ou seja, que me pedem para fazer ou as sugestões de trabalhos que eu dou, que sempre sejam com dados. Então, às vezes estou a fazer outras coisas, que faço com muito gosto, mas não é com dados, e sempre que posso tento pelo menos ter uma base factual de estatística. Verdadeiramente, desde que eu entrei no Expresso, eu não fiz muitos trabalhos de jornalismo de dados, por todas essas coisas que eu disse, ou seja, porque é difícil, porque leva tempo, precisaria de uma equipe dedicada, e não há uma equipe de jornalismo de dados. Nós trabalhamos com o departamento de infografia. São coisas que partem da iniciativa de uns ou outros, mas sempre com essa ligação entre o jornalista e a infografia. Há uma coisa que tenho pena aqui, e lembro que no Público funcionou muito bem, que era ficávamos sentados quase ao lado uns dos outros, a programação funcionava ao pé de nós, o programador estava sentado a 2 metros de mim. Nós almoçávamos juntos e às vezes as ideias surgiam no almoço e depois concretizávamos. Isso é que eu acho que é uma falha, e ocorre não é só no jornalismo de dados, mas no multimédia também. Raquel a nossa equipe de programação está totalmente desassociada da redação, não apenas fisicamente, mas também nas funções pois eles estão por conta da parte técnica de aplicações e websites, não só do Expresso, mas dos outros veículos do grupo como a SIC, ou por exemplo os audiolivros. Estamos a utilizar o Tableau, por exemplo, que é uma ferramenta que tem muito mais limitações do que se fizéssemos uma coisa de raiz.

**Sofia** - Gostávamos de fazer alguma coisa mais original, mais nossa, mas normalmente eu só estou mesmo com a equipe dos programadores quando eu tenho um bloqueio ou quando percebo que preciso de ajuda, mas não na questão habitual do conteúdo.

#### **4. Qual o tamanho da equipe? A equipe é dedicada somente ao JD?**

**Raquel** - Somente eu, e a infografia que trabalha para jornal inteiro. Eu trabalho na secção de Sociedade e faço reportagens do dia-a-dia. Dentro de sociedade está saúde, educação, justiça, ciência e uma série de outras coisas, portanto, nesta secção realmente cabe muito mais temas onde é possível trabalhar com dados e acaba de por ser por isso que eu estou nesta secção, mas é claro que se for necessário fazer algum trabalho de dados para política, por exemplo, eu faço.

**Sofia** - eu sou da secção de infografia que acaba por ser mais transversal. O trabalho da infografia está normalmente inserido em todas as secções, principalmente sociedade política, economia...

#### **5. Existe uma editoria ou sessão separada no site para dados/infográfico?**

Não.

#### **6. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

**Raquel** – Depende. Se há uma semana em que eu acabei o meu trabalho mais cedo então posso me dedicar a procurar dados para uma reportagem. Vou aos sites de organismos públicos, de dados abertos, de vez em quando consigo fazer isso, mas, se calhar, na semana seguinte alguém me pede um trabalho com dados ou alguém fala de um relatório que saiu.

**Sofia** – E também há determinados eventos, por exemplo, o ranking das escolas, eleições, que mesmo antes dos dados estarem disponíveis já temos um pré-conhecimento. Então já pensamos sobre o que vamos fazer, se vamos replicar o que fizemos nos outros anos ou vamos fazer algo novo. Há esse planeamento.

#### **7. Existe algum treinamento para a equipe (exemplo: coding para os jornalistas, jornalismo para programadores)?**

**Raquel** - Eu faço a pós-graduação de Visualização de Informação na Nova FSCH. O jornal teve essa visão de querer investir, no caso, através de uma pós-graduação. O que aprendi antes foi por minha conta.

**8. Vocês acabaram adquirindo novas habilidades, aprendendo a usar ferramentas?**

**Quais as ferramentas/softwarewares que vocês usam?**

**Sofia** - Na minha área as coisas mudam muito em relação à tecnologia e o fato também de eu trabalhar com várias secções, seja para o papel, seja para o online, me obriga a ter conhecimentos técnicos de acordo com o trabalho que estou a fazer. O que eu sinto também é que eu acabo por não dominar profundamente um programa ou uma área, acabo por saber um bocadinho de código que me permite resolver algumas questões quando uso o Tableau ou outras ferramentas open source, que tem as ideias de serem acessíveis a pessoas sem conhecimento de programação, mas também não são assim tão imediatas nem tão fáceis de trabalhar, principalmente quando queremos algo mais customizado. Então sei um bocadinho de programação. O D3.js por exemplo já não consigo usar. Sei um pouco de HTML e CSS. No papel usamos muito o Indesign e o Illustrator. Também tem que ter alguns conhecimentos de Photoshop e After Effects para vídeos que tenham infografia. Usamos o Tableau para fazer alguma coisa de interatividade, mas é limitado e de todo não funciona bem numa das plataformas que mais crescem, que são os telemóveis. Usamos a versão gratuita e não conseguimos fugir muito dos modelos que eles têm, não dá para customizar muito, só até certo ponto, mas não há liberdade total. Houve uma altura que usamos também o Animator da Adobe.

**Raquel** - Eu sabia trabalhar com Excel minimamente, mas as tabelas dinâmicas por exemplo... foram uma coisa que, de facto, a partir do momento em que você sabe trabalhar com isso, consegue fazer tudo. Ou seja, eu não sou especialista em Excel, mas consigo utilizar ao ponto de saber procurar determinada função para fazer determinada coisa. Outras ferramentas como Google Refine, o Tableau... Na verdade, eu aprendi a fazer o básico no Tableau e resolver problemas. Criar coisas novas no Tableau eu não consigo e nem sequer é uma ferramenta que eu uso regularmente. Eu fui sentindo a necessidade de começar a perceber cada vez mais de programação. Agora na pós-graduação tenho uma cadeira de Python estou a adorar. A pós-graduação eu faço pelo jornal, mas os outros que

eu fiz anteriormente foram por minha conta. Temos cadeiras de análise de dados, de estatística, e eu sinto cada vez mais que me ajuda saber sobre estatística.

**9. Do ponto de vista da recepção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

**Sofia** - Estou aqui a pensar numa reportagem que eu fiz sobre os ataques terroristas no mundo. Não existe uma base de dados, então o que eu fiz foi pegar as várias que existiam e fazer o cruzamento e depois pesquisar pelas agências de notícia e cruzar com outras notícias e fazer um mapeamento desses dados. Eu recebi várias mensagens de leitores a perguntar e a pedir as bases de onde eu tinha conseguido os dados, ou seja, eu acho que quando se faz um trabalho de investigação há retorno do leitor. Mas eu acabei por abandonar isso porque atualizar não é compatível com trabalho diário, é preciso ter alguém para fazer só isso e eu não posso manter. Mas eu sinto em grande parte das situações que as pessoas reconhecem o valor desses trabalhos.

**Raquel** - Todos esses trabalhos, principalmente os interativos, eu acho que os leitores, independentemente de serem mais escolarizados ou não, dão valor a este tipo de coisa.

**10. Quais os desafios para a prática de jornalismo de dados atualmente?**

**Sofia** - A questão do acesso aos dados. Muitas vezes eu fico a me perguntar se não somos nós que nos autolimitamos, ou seja, se bastava pedir acesso a esses dados em bruto. Pode não ter interesse para um público geral, mas podem estar disponíveis nas empresas ou nos organismos. Temos que começar a nos desafiar mais a nós próprios e tentar aceder a esses dados.

**Raquel** - Acho que há um lado que é nosso, mas acho que se deve muito a quantidade de coisas que aparecem, a velocidade em que se pedem as coisas, e há também o lado da própria desconfiança dos sítios quando nós pedimos os dados. Há muitos dados que deveriam estar públicos, mas não estão porque há um receio muito grande de qual vai ser a utilização que poderíamos fazer. O outro desafio é, e eu penso que tem que ser eu a fazer isso, ter sempre um trabalho a andar a médio prazo, temos que aceitar que tem que ser assim, que vais ter sempre trabalhos do dia-a-dia e mais um trabalho que não tenha prazo, que dê para ir fazendo. Eu até acho que com o Tableau nós até conseguimos saltar a parte de as visualizações não ficarem totalmente estáticas, mas é verdade que tem muitas

limitações e nós gostávamos de fazer as coisas da nossa maneira, montar de raiz uma coisa que nós pudéssemos definir.

**Sofia** - É nessa parte que faz falta o tal do programador. Seria simples, e é possível. Com três ou quatro pessoas faz-se um trabalho enorme. Agora em Pamplona eu ouvi uma conferência do editor da Reuters em que ele dizia que eles têm 7 pessoas, ou seja, não é uma equipe tão grande, mas todos sabem um bocadinho de cada coisa, todos têm noção de programação, mas há um ou dois que são realmente muito bons a fazer programação.

**Raquel** - Eu acho que se eu nunca tivesse tido a experiência de conseguir fazer como fiz na época do Público, eu própria questionaria se não seria preciso ser um sítio maior e com mais investimento para de fato concretizar isso. Mas ainda bem que em determinado momento eu consegui fazer isso, pois por um lado eu fiquei dedicada ao projeto por não ser funcionária do jornal, mas também tinha muito a ver com a forma como Público estava organizado. Estavam as peças certas no sítio certo. Havia pessoas com todos os conhecimentos no mesmo sítio, sentadas mesmo uma ao lado da outra, com esses conhecimentos de programação, de design, de estatística. Portanto estava tudo montado para que isso fosse possível. É preciso ter 3 ou 4 pessoas com diferentes valências de jornalismo de dados, e isso em qualquer sítio. É preciso uma pessoa de cada área com as valências necessárias e com tempo para fazer as coisas.

#### **11. Qual a disponibilidade da empresa para investir mais em JD?**

**Raquel** - Eu sinceramente acho que disposição até existe para criar um núcleo de jornalismo de dados, mas na prática é claro que as coisas são mais difíceis todos os meios de comunicação.

**Sofia** - Abertura e vontade acredito que tem. Eu acho que já ouvi alguma conversa no sentido de incluir um programador, mas nunca senti que isso fosse realmente acontecer.

### **VIII. ENTREVISTA COM PEDRO PIMENTEL – JORNAL DE NOTÍCIAS**

Peço imensa desculpa pela demora na resposta, mas depois de insistência com a direção e editorias a ver se poderíamos responder de melhor forma, concluímos que não poderíamos ser mais assertivos.

Eis a resposta do departamento de infografia que maior tem ligação àquilo que pode ser designado de jornalismo de dados.

Passo a citar:

“Não posso responder a todas as perguntas porque acho que o Jornal de Notícias não tem propriamente jornalismo de dados no âmbito pleno que o inquérito refere. Ficam apenas alguns dados sobre a Secção de Infografia

**1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

No Jornal de Notícias não existe especificamente jornalismo de dados.

Existe uma Secção de Infografia que entre outras competências trata também diferentes tipos de dados (económicos, sociais, desportivos, eleitorais, etc.)

A secção é coordenada por um jornalista/infografista, Isidro Costa Esteves, 53 anos, que trabalha no jornal desde 1985. Inicialmente estava integrado na Secretaria de Redação, onde tratava os dados relacionados com os mercados bolsistas.

A secção de Infografia foi criada em 1991, chegou a ter três jornalistas/infografistas.

Atualmente tem a colaboração de um infografista/ilustrador, João Correia

**3. Existe uma editoria ou sessão separada para dados/infográfico?**

Sim. A secção de Infografia

**11. Qual a disponibilidade da empresa para investir mais em JD?**

A empresa não tem previsões a curto prazo para um maior investimento no jornalismo de dados”

**IX. ENTREVISTA COM RITA MARQUES COSTA – PÚBLICO**

**1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Rita Costa, 24 anos, jornalista de dados. Eu estudei Ciência da Comunicação na Nova, estagiei no Expresso, depois estive em Londres por um período pois achava que não conseguiria emprego em Portugal, mas não foi nada a ver com Jornalismo, regressei e trabalhei numa revista muito pequenina sobre agricultura, e foi aí que, no fundo, comecei a ter mais interesse pelo Jornalismo de Dados e a perceber o que eu poderia fazer para sair dali e conseguir me diferenciar das pessoas e fazer coisas mais interessantes que não fosse só, ok, é muito importante contar histórias, mas fazer coisas que me destacassem das outras pessoas. Já gostava muito de trabalhar no Excel e mexer com números, então pensei: vou aprender isso. Comecei a fazer cursos online. O primeiro foi sobre visual storytelling com Alberto Cairo no Night Lab, ele é incrível. E daí, comecei a me interessar, fazer mais cursos, fiz um curso de Excel básico e depois fiz um curso de comunicação visual da informação no ISCTE, onde ganhei mais sensibilidade para esses temas. Depois percebi que eu precisava aprender um bocadinho a programar. Fiz o curso de python para jornalistas com o Ben Welsh, também do Knight Center, claro que não sei programar em python, mas dá uma base e sensibilidade. Isso foi tudo num período de dois anos, e veio a ideia do prêmio das indústrias criativas. Eu pensei: não há trabalho nesta área em Portugal, então vou fazer o meu próprio trabalho. Com um outro amigo, pensei em me candidatar ao prêmio com a ideia de fazer jornalismo de dados para os jornais portugueses. Fomos à final, mas, entretanto, o Público pôs um anúncio para selecionar um jornalista de dados e eu candidatei-me. E pensei bem... achei que seria bom, seria um outro meio para chegar ao mesmo fim. Então nem pensei duas vezes.

No Público, a ideia inicial era trabalhar no projeto Cidades, que é um projeto de jornalismo local, com o objetivo de aproximar as pessoas do país inteiro do jornalismo local. Basicamente, é uma plataforma onde a pessoa pode pesquisar o seu conselho e recebe notícias do local. Também temos dados estatísticos de cada conselho, apresentados de uma forma mais gráfica. A ideia era fazermos mais histórias sobre os conselhos de Portugal. Eu fiz uma sobre casamentos homossexuais, e em metade do país nunca houve casamento de pessoas do mesmo sexo. Teve a questão da imigração nos diferentes conselhos e distritos... este tipo de trabalhos. Entretanto, eles perceberam que era preciso fazer mais coisas do que isso, e eu mudei para a secção de Sociedade, onde faço muito trabalho de

dados em Portugal inteiro. E vou trabalhando também com outras secções, pois é inevitável.

## **2. Breve histórico do JD no veículo. Iniciativa de quem?**

Eu acho que a Raquel Albuquerque, que está no Expresso, foi a primeira pessoa a fazer Jornalismo de Dados no Público. Mas também sei que sempre houve sensibilidade para este tema, se calhar não nomeadamente como Jornalismo de Dados, mas sempre houve um reconhecimento da importância dos dados no jornalismo. O Público foi o jornal que possibilitou que se fizessem o ranking das escolas. O Ministério não dava os dados sobre as médias dos exames nacionais por escola, e foi o Público que insistiu muito, muito, muito, até que agora eles libertam esses dados e toda a gente faz essa análise. Formalmente, se calhar, foi coma Raquel mesmo que se reconheceu que se fazia Jornalismo de Dados. Também acho que tinha lá outra pessoa que se interessava pelo tema, e sei que iniciativa foi mesmo muito dessas pessoas.

## **3. Existe aqui um grupo ou um contato entre os profissionais de Jornalismo de dados?**

Não. Por acaso, é uma coisa que acho que deveríamos fazer, e eu queria muito tentar criar algumas sessões em que pudéssemos todos partilhar ideias e aquilo que fazemos. Eu sinto imensa necessidade. É muito difícil, em uma redação, as pessoas terem a perceção daquilo que estou a fazer, e às vezes as pessoas não fazem ideia do que eu estou a falar a fazer, e sinto a necessidade de ir ter com outra pessoa que perceba aquilo que estou a dizer. Às vezes na redação, alguém diz: pois, se achas que é assim, é assim. Há muitas pessoas que não sabem mexer no Excel.

## **4. Existe uma editoria ou sessão separada para dados/infográfico?**

Tem a secção multimédia e, dentro, tem infografia, mas os meus trabalhos quase nunca vão para essa secção. Tem um só que eu fiz, com a ajuda da infografia. É uma secção do departamento de infografia, o que não significa que os outros trabalhos não tenham infografia. Quando há texto, vai sempre para a secção específica. Eu lembro-me que fiz uma sobre o poder de compra nos municípios, e tinha uma infografia muito fixe, interativa, às pessoas podiam clicar. Esse foi para a secção específica, no caso, local.

## **5. Qual o tamanho da equipe?**



Oficialmente, como jornalista de dados, sou somente eu, mas eu trabalho também com o departamento de infografia, mas eles só fazem o design, não trabalham os dados. E às vezes com o webdesign. Oficialmente, não há programadores. Quando é preciso, o webdesign faz alguma programação. Outro exemplo: os dados do ranking das escolas vêm em uma base de dados do Access, uma coisa gigantesca, que nem eu sei mexer, ou seja, para transformar aquilo em uma coisa inteligível temos a ajuda dos informáticos, que também trabalham connosco quando necessário. Eles são mesmo do departamento de informática, da parte técnica, não estão ligados à redação, mas sempre ajudam.

#### **6. A equipe, no caso você, é dedicada somente ao JD?**

Não. Eu faço outras coisas. Eu tento sempre dar esse "twist" para o lado dos dados, mas às vezes não dá, e eu faço outras coisas.

#### **7. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

Depende. Há muita coisa que fazemos por sugestão dos editores. Temos muito trabalho para o dia, pois o jornal é diário, então muitas vezes surgem coisas que não têm nada a ver com dados para o dia. E muita coisa que tem a ver com dados sou eu que sugiro, porque tenho mais sensibilidade para isso. Mas editores também às vezes sugerem.

Há diferentes formas de chegar à ideia. Por exemplo, na pauta das ruas com nomes de mulheres, pelo dia da mulher, eu sabia que eu queria fazer algo com os nomes das ruas, mas tinha o problema de não haver uma base de dados com todos os nomes de ruas em Portugal. Então o problema foi chegar aos dados. Eu encontrei a Central de Dados, que é uma fonte de informação, que não é minha fonte favorita pois eu sei que eles não são a fonte oficial. Então eu tento sempre dar esse contexto ao leitor, para que ele perceba que isso já foi tirado por outra pessoa. Mas temos fontes que são mais oficiais. Temos o portal da transparência, que são os dados da saúde, temos o dados.gov, que é o portal de dados abertos do governo, temos o Lisboa aberta. Uma dificuldade gigante em Portugal é que não há muita coisa de dados abertos. Isso é uma falha gigante. Mesmo nesse portal da transparência nos encontramos erros absurdos. Por exemplo, legendas das definições dos dados erradas, dados desatualizados. Temos que ter muita atenção. A minha fonte oficial favorita é o INE, Instituto Nacional de Estatística. Eles são muito bons e têm muita atenção

ao detalhe. Nunca encontrei erro nos dados deles. Eu diria que, para cada secção, até há muita coisa, mas os dados mais interessantes não existem. Nós também fazemos webscrapping, mas é difícil fazer, as ferramentas são difíceis de usar. Nós tentamos sempre usar os dados semi-estruturados. Sempre fazemos algum tipo de análise

**8. Existe uma periodicidade definida para a publicação de “data stories”? Se sim, qual e por quê?**

Não há. Normalmente, se existe alguma efeméride, por exemplo, o dia internacional da floresta, da água, tentamos fazer coisas com dados. Como foi o dia da mulher. Quando houve os fogos, tentamos contextualizar muito com dados de anos anteriores. com as cheias também.

**9. Quais as ferramentas/software utilizados para produzir “data stories”?**

Eu uso muito o Excel. Estou a aprender Python e R, pois sei que vai me ajudar bastante para fazer uma análise mais rápida, cuidada e assertiva dos dados. Para visualização, as pessoas da infografia usam bastante o Illustrator e para as interativas eles usam .... Sei que as pessoas do webdesign às vezes usam o D3.js. Eu uso muito o Tableau, até para fazer a análise dos dados, pois é fixe ver logo em vez de estar a fazer os gráficos no Excel. Nos mapas, uso muito o Arcgis.

**10. Existe algum treinamento para a equipe (exemplo: *coding* para os jornalistas, jornalismo para programadores)?**

Há duas pessoas que estão a fazer a pós-graduação em visualização de informação na Nova. A Cláudia Carvalho, jornalista do online, e a Célia Rodrigues, da infografia. Eu faço um mestrado em Análise e Gestão da Informação, na Nova IMS. É muito técnico, mas está a ser incrível pois ajuda a perceber o que está por trás.

**11. Do ponto de vista da receção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

Eu acho que depende muito dos temas. A pessoa, quando clica no título, não sabe se aquilo é um gráfico interativo ou não. Eu sei que a das mulheres gerou burburinho, mas acho que foi pelo tempo. Acho até que é uma boa ideia criar uma área só de gráficos. O que eu vejo é que, em Portugal, o leitor ainda não tem a literacia para ler gráficos.

## **12. Quais os desafios para a prática de jornalismo de dados em Portugal atualmente?**

Uma dificuldade gigante em Portugal é que não há muita coisa de dados abertos. Isso é uma falha gigante. E, mesmo no portal da transparência, nós encontramos erros absurdos. Por exemplo, legendas das definições dos dados erradas, dados desatualizados. Temos que ter muita atenção. E o maior problema aqui em Portugal nem é o fato de não se disponibilizarem dados. O grande problema é que os dados não existem. Não se medem. Por exemplo: não temos dados sobre raça. Agora é que, se calhar, no censo de 2020 vão colocar uma pergunta sobre raça.

## **13. Qual a disponibilidade da empresa para investir mais em JD?**

Acho que existe uma disponibilidade, sim. Eu entrei no meio do ano passado, é recente. Eles identificaram isto como algo importante e percebem que isso é o futuro. Quando vais ver um artigo online, isso chama muito mais a atenção do que um texto corrido, e eles percebem isso e que há muitas histórias ainda por contar porque não olhamos direito para os dados. E eu estou a aprender mais sobre isso no mestrado e, como eu disse, há mais duas pessoas, como eu disse, a aprender mais sobre isso na pós-graduação.

## **X. ENTREVISTA COM LEO XAVIER – OBSERVADOR**

### **1. Nome, idade, cargo e breve descrição da carreira até chegar ao jornalismo de dados.**

Léo Xavier, 34 anos eu sou designer e programador, a minha formação de facto é em belas-artes, em design. Eu sou dois em um. Eu estou no jornal desde o lançamento, sou diretor técnico do jornal. Eu dirijo a equipa de design e tecnologia. E, por outro lado, há dois anos nos candidatamos ao funding do Digital News Initiative da Google, que procurava projetos de inovação na área de jornalismo online. Então, é deste projeto que surgiu a Frames. Nós nos candidatamos como o jornal Observador, a Google forneceu os fundos para o projeto e, em vez de fazer dessa ideia uma coisa interna e exclusiva para o Observador, decidimos fazer uma spin-off, uma empresa separada coparticipada pelo Observador e por mim. A Frames está a montar este produto que já está a ser usado pelo Observador e agora estamos a exportar para outros publishers em outros países um produto para melhorar os

artigos através da comunicação de gráficos contextuais. Portanto eu ainda sou diretor técnico parte do tempo, e a outra parte estou então a liderar a Frames.

## **2. Então essa parte de jornalismo de dados do Observador é uma parceria com a Frames?**

Mais ou menos. Nós temos uma equipa própria de jornalistas na Frames que no dia a dia olha para os principais eventos da atualidade, decide quais são os temas mais relevantes, pesquisa dados sobre eles e prepara pequenos gráficos, coisas menores tanto na parte visual quanto no tamanho ocupado, que deem informação sobre o tema. Se aconteceu, por exemplo, um terremoto no México, pode ser um histórico de terremotos recentes ou uma explicação da escala Richter. Se for taxa de desemprego, obviamente a taxa histórica de desemprego em Portugal e assim por diante. Esses frames são depois colocados automaticamente pelo nosso sistema nos artigos, ou seja, sempre que o Observador publica um artigo, nosso sistema analisa automaticamente o seu conteúdo, classifica-o e procura na base de dados da Frames se há alguma coisa que pode ou não fazer sentido. O jornalista pode remover ou não, ou pode até criar o seu próprio, mas na maioria das vezes o jornalista não sabe, não tem tempo ou não tem paciência, nem informação, nem ferramentas. A decisão é sempre do jornalista, mas se ele não fizer nada aparece automaticamente, ou seja, nós presumimos que isto está certo e o jornalista se quiser ele tem que ativamente intervir e remover. Isso porque sabíamos, por experiência própria, que se deixarmos nas mãos dos jornalistas todo este trabalho de pesquisa de dados e inclusão do gráfico no artigo, nós normalmente teríamos uma cobertura de artigos com gráficos menor do que um por cento. Durante 3 meses no ano passado, foram 0,34% de artigos com gráficos quando a única ferramenta era o Infogram. Com esse novo sistema, no mês de fevereiro fizemos 54% de artigos com gráficos, o que é inexistente no mundo, não há jornal que tenha tantos gráficos nos seus artigos.

Todo esse esforço da Frames é para esses gráficos pequenos e contextuais, mas tudo o que são trabalhos mais a fundo de visualizações mais complexas são coisas que a Frames não tem nada a ver e obviamente que cada jornal produz na mesma. No caso do Observador, é a minha equipa que os faz, e são coisas como as eleições ou uma infografia qualquer sobre um tema mais complexo. Não fazemos muita coisa, mas quando fazemos obviamente são produções maiores. Portanto temos duas frentes: a automatizada contextual com coisas

pequenas e aquilo que, vamos chamar assim, são megaproduções e neste caso inclui interatividade.

### **3. Breve histórico do JD no veículo. Iniciativa de quem?**

O projeto da Frames começou a ser desenvolvido há um ano e meio e a ser publicado há um ano. Os projetos maiores existem desde sempre no Observador, que foi lançado há três anos e meio, quase 4 anos, com a ideia de ser um jornal unicamente digital, uma inovação. Logo no início quando lançamos havia as eleições e portanto, já fizemos projetos interativos com os resultados eleitorais.

### **4. Existe uma editoria ou sessão separada para dados/infográfico?**

Secção específica não. Temos conteúdos interativos que são destacados nas respectivas secções temáticas com um prefixo azul a dizer que é interativo.

### **5. Qual o tamanho da equipe? A equipe é dedicada somente ao JD? Quem são os profissionais e quais as formações deles?**

Não há uma equipe específica para isso. Tipicamente, quando há uma produção maior interativa, quem faz sou eu o meu senior designer, e depois depende do esforço de programação, que pode ser de um, dois ou três programadores. Depende. Sempre tem um editor da secção específica do tema e programadores que não tem nenhuma skill de jornalismo, mas trabalhamos muito perto da redação partilhamos o mesmo espaço, portanto estão habituados a lidar com esses temas. De vez em quando algum jornalista pode fazer um bocadinho de limpar dados ou raspar dados, mas tipicamente na nossa redação não tem aquilo que eu chamaria de um data journalist.

Na Frames há, inclusive o jornalista tem um phd em data visualization. Achamos importante ser uma curadoria editorial. Nesse caso é um jornalista a tempo inteiro para Portugal que recolhe dados e decide temas, mas isso ocorre completamente fora do contexto do Observador.

### **6. Como são definidas as pautas das “data stories”? Descreva as fases do fluxo de trabalho na produção de uma “data story”.**

No Observador, quando a redação tem uma necessidade, e são coisas grandes como as eleições, os mundiais, probabilidades de uma equipa de ganhar um campeonato,

normalmente há um pedido e nós rapidamente respondemos se é possível ou não, por uma questão de tempo e não de skills. Nós fazemos tudo: pensamos onde buscar os dados, na implantação técnica, na interatividade, no design e é sempre tudo mobile-first. Nós somos muito autônomos, obviamente estamos sempre consultando a redação, mas eles confiam muito em nós, os designers e programadores. A intervenção jornalística é mesmo majoritariamente na sugestão. Isso por conta dos muitos anos que estamos na equipa e eu próprio já trabalhei com outros jornais, portanto é uma área que eu sempre tive interesse. No Observador eu acabo por ter um papel que seria de product manager, pois no fundo estamos a desenhar um produto. A redação tem a ideia e as necessidades, a equipe comercial tem objetivos, depois junta-se tudo na materialização do produto digital. E por sermos só digitais, isso então é mais direto do que no offline.

**7. Existe uma periodicidade definida para a publicação de “data stories”? Se sim, qual e por quê?**

No Observador, eu diria que só fazemos interativos com data visualizations duas ou três vezes por ano. É muito baixa a cadência, porque o que fazemos é mais histórias visuais, aqueles long forms multimídia.

Na Frames, não há uma periodicidade definida. Há dias que temos um ou dois, outros temos três ou quatro. Muito tempo é despendido em manter a informação atualizada porque cada frame criado é continuamente atualizado. Por exemplo, a estatística de emprego quando sai não vamos criar um novo frame, nós atualizamos o antigo. Então hoje temos um acervo por volta de 500 ou 600 frames já criados e continuamente atualizados. Apesar de cada dia criarmos dois ou três novos, a verdade é que já temos uma massa crítica que nos permite chegar aos tais 54% de artigos com gráficos no Observador.

**8. Quais as ferramentas/softwarewares utilizados para produzir “data stories”?**

No Observador, os jornalistas usam bastante o Infogram para criar coisas próprias, mas como eu disse é uma coisa reduzida, situações mais pontuais. Eles têm essa liberdade e até há um incentivo para fazerem isso. Quando são as produções interativas maiores, não há nenhuma ferramenta que usamos. É programação mesmo nossa, código nosso, direto. Numa situação outra talvez tenhamos usado o D3.js. O nosso backoffice é Wordpress, portanto a linguagem é PHP, mas o grande trabalho é de java script, é a linguagem principal.

Na Frames temos nosso próprio sistema para construir gráficos, e usamos o Node, um environment de JavaScript.

**9. Existe algum treinamento para a equipe (exemplo: *coding* para os jornalistas, jornalismo para programadores)?**

Treinamento tecnológico para os jornalistas, no início tínhamos um bocado dessa visão, mas não pegou. A malta é muito resistente e para usar o infogram não é preciso perceber nada de código. O incentivo é mesmo editorial. Os diretores do jornal e os editores fazem pressão nos jornalistas para fazerem mais gráficos com Infogram.

**10. Do ponto de vista da recepção, qual o impacto das “data stories” para o leitor? O jornalismo de dados tem criado mais valor ao produto jornalístico da empresa?**

Sim, gera page views, mas normalmente nem colocamos publicidade nesses interativos pois são reportagens a fundo, exclusivas, são acima de tudo maneiras de posicionar o jornal como conteúdo de qualidade, e depois capitalizar nisso com receitas indiretas.

**11. Quais os desafios para a prática de jornalismo de dados atualmente?**

Eu acho que o mais difícil, eu diria é ter os recursos, ou seja, alguém da redação ter os skills, e para ter skills é preciso ter tempo e recursos. No fim, é tudo a mesma coisa: recursos. O tempo que leva a fazer é de 10 artigos de texto versus um artigo de dados, e a proporção é até pior que é essa. Então é preciso ter um jornalista que tem conhecimento mínimo técnico de dados isso é muito raro.

## **XI. ENTREVISTA COM LUIS VARGAS – HEMICICLO**

**1. O que é o Hemiciclo?**

É um projeto de iniciativa pessoal com o objetivo muito mais de informar do que de produzir opinião. A ideia é ser sempre o mais neutro e analítico possível.

**2. Como surgiu a ideia?**

Nós pensamos em fazer o site porque há uns anos houve uma iniciativa em Portugal que resultou numa proposta apresentada à Assembleia da República no sentido de transparência parlamentar. Uma parte dessa proposta era que a Assembleia fornecesse esse tipo de informação, mas ela não foi aprovada e, portanto, a única maneira era que

alguém externamente, ou seja, da sociedade civil, conseguisse montar a coisa e eu como programador achei um desafio bastante interessante. Na altura, falei da ideia com o David, que eu já conhecia e foi a pessoa ideal para fazer isso pois ele tem conhecimento profundo do funcionamento da assembleia e do processo legislativo em Portugal. Enquanto eu programava, e fazia a parte do código, foi fundamental ter alguém como ele a orientar a arquitetura da coisa. Ele tem formação em Economia, mas tem experiência em Política e processo legislativo porque trabalhou para o Parlamento Europeu. A minha formação é em Design Industrial, mas atualmente sou programador, faço a parte de programação e design do front-end de web sites.

### **3. O site tem algum retorno financeiro?**

Não, isso não tem retorno nenhum. Já pensamos em colocar publicidade para nos ajudar com os custos do servidor, que também não são nada de especial, mas também não queríamos encher o site com publicidade, portanto, de momento, está mesmo assim.

### **4. E também acaba perdendo um pouco da credibilidade, não é?**

Sim, ou seja, são cerca de 15 € por mês, não é assim nada de muito. Obviamente que quando mencionam o site na televisão o servidor vai abaixo, pois não tem grande capacidade para tanta carga simultânea.

### **5. Quais são as ferramentas e a tecnologia que você utiliza para fazer o site?**

A parte mais complicada é curiosamente aquela que não se vê, porque toda a informação que nós temos não é produzida por nós, ela é raspada automaticamente por um script, ou seja, um robô que duas vezes por dia visita o site da Assembleia da República e raspa milhares de páginas para extrair a informação. Esse robô também organiza a informação na nossa base de dados. Fazer essas raspagens foi a parte mais difícil, porque o site da Assembleia está feito numa coisa chamada ASPX, que é uma tecnologia do início do século, e para raspar era preciso fazer pesquisas, inserir coisas em campos, guardar cookies e ir saltando de página em página. A navegação automática nessas páginas não é trivial. Uma coisa é nós sabermos já o endereço das páginas que nós queremos raspar, mas muitas vezes nós temos que entrar numa página e, para aceder à página seguinte, temos que preencher um formulário. Então a ideia foi fazer um script que conseguisse preencher os formulários, submeter os formulários e guardar os cookies gerados por esses formulários.



É a maneira que o site tem de perceber o nosso histórico e, portanto, navegar automaticamente nas páginas, que são públicas.

Depois, o próximo desafio foi estruturar a informação. Neste momento, temos só três legislaturas, mas a ideia é trabalhar para ficar com nove e, portanto, recuar até os anos 90. Mas, mesmo só com as três legislaturas (desde 2009), é muita informação. Quando o utilizador entra no site e faz pedidos à base de dados, para aceder a essa informação convém que ela esteja estruturada de uma maneira que o site não fique muito lento. Então começamos a usar uma base de dados, que é uma coisa relativamente recente, “no SQL”, feita por objetos. O nome dela é Mongo DB. Uma base de dados clássica, SQL, funciona por tabela, mais ou menos como um ficheiro Excel, com linhas e colunas. A Mongo DB tem o funcionamento completamente diferente, ou seja, cada diploma que nós temos na nossa base de dados é uma estrutura parecida com uma árvore, com vários ramos, e cada ramo pode ter sub-ramos ou coleções de outras árvores. Foi a única maneira que nos permitiu por isto a funcionar com tempos de acesso razoáveis. Pretendia-se que quando uma pessoa clicasse no site, o pedido demorasse menos de meio segundo. Para isso, nós aqui estamos a fazer algumas funções com uma matemática mais intensiva. Se fosse com uma base de dados por tabela, demoraria vários segundos.

Tendo isso resolvido, depois foi uma questão de fazer a interface, a visualização. Esse é um processo que ainda estamos sempre a acrescentar mais coisas. Há pouco tempo, acrescentamos o índice de concordância.

## **6. E o que você utiliza para construir a visualização?**

Utilizamos algumas vezes o D3.js., por exemplo, na parte das votações, quando a visualização do hemiciclo propriamente dito tem as bolinhas coloridas e podemos ver aprovação, abstenção e rejeição. O resto é JavaScript, basicamente. Usamos o Node, que roda o JavaScript. O site é uma coisa que é 100% java script, quer do lado do utilizador, quer do lado do servidor. Portanto, do lado do servidor temos essa base de dados, a Mongo DB, que é uma coisa que funciona em java script. A parte toda do back end, do processamento e do algoritmo é feita em Node. E depois, na parte do front end é tudo java script e HTML. Era uma coisa tão específica que era mesmo preciso sermos nós a desenvolver a coisa toda customizada. Por coincidência, eu também estava a fazer muita

coisa com scraping e com data visualization e, então eu estava mesmo com vontade de pegar esse desafio para ver o que eu conseguia fazer.

**7. E essa questão da visualização também está bem relacionada com a sua área de designer industrial, não é?**

Sim. É uma área que me interessa bastante e é uma área de futuro. Cada vez recolhemos mais dados de mais sítios e, portanto, não é só difícil processar os dados todos, mas também visualizá-los. Então é uma área de futuro.

**8. A sua formação como programador é autodidata?**

Sim, completamente. Eu trabalhei numa empresa onde fazia design industrial, a YDreams. Lá, a ideia sempre foi fazer coisas interativas, então eu sempre tive muito contato com programadores. Depois, quando eu saí, não foi propriamente uma decisão de carreira, mas era uma coisa que me tinha fascinado e comecei a aprender.

**9. E hoje você trabalha como autónomo?**

Sim. E eu tenho vários projetos como o Hemiciclo, que é um projeto pessoal paralelo. Alguns mapas, por exemplo.

**10. E você tem um blog para divulgar isso?**

Por acaso não. Eu devia tratar disso mesmo.

**11. Como tem sido o retorno de acessos? Você disse que quando passa na televisão o servidor nem aguenta, mas e no geral?**

O lançamento foi espetacular, excedeu largamente as nossas expectativas. Foi em setembro de 2017 e o servidor estava sempre indo abaixo, porque era muita gente a aceder. Desde então, os acessos foram crescendo e tem tido alguns picos. Saíram notícias no Diário de Notícias e no Público recentemente, porque foi incluído o índice de concordância, então tivemos muito mais tráfego. Às vezes, quando há votações que as pessoas já sabem que vão ocorrer porque saiu nas notícias durante a semana, nós temos muito mais tráfego que o normal. Estamos agora a pensar um esquema para atrair mais pessoas. Temos perfis no Facebook e no Twitter e atualizamos essas contas pelo menos umas duas vezes por semana. Uma coisa que nos interessava era estabelecer uma parceria com um jornal para mais tráfego. Poderia ser um esquema de troca com um órgão de

comunicação social, em que nós fornecêssemos conteúdo, ou eles até nos poderiam pedir, e nós faríamos visualizações. Em troca, eles publicariam ou linkariam para o nosso site.

**12. Há uma sessão de notícias chamada Atualidades. Quem faz a atualização e qual a periodicidade?**

Somos os dois, quando temos tempo, porque nós também temos muitas coisas para fazer. Vamos pegando nisso mais como um passatempo.

**13. E o resto do site tem algum tipo de manutenção?**

Sempre que tem votações, geralmente às sextas-feiras, nós temos que fazer alguns ajustes. É uma coisa que nos consome às vezes uma hora, mas é só quando há votações desalinhadas, ou seja, quando há deputados que votam de maneira diferente da sua bancada.

**14. Vocês têm mais algum plano para o site?**

Temos um plano muito ambicioso. Daqui a um mês, vamos ter uma sessão com podcast, que vai ser feita em parceria com um grupo especializado. Vamos começar a ter em entrevistas com pessoas da Assembleia da República. Além disso, estamos a recuar até os anos 90 para termos muito mais diplomas e, depois, vamos ter uma parte que é importantíssima e é muito difícil de fazer que é a secção das comissões parlamentares, pois grande parte do trabalho na assembleia nem é tanto no plenário. O nosso objetivo é a montar um projeto e concorrer a financiamento para conseguirmos estar dedicados, se não exclusivamente, quase exclusivamente a cumprir todas essas etapas. Queremos que as pessoas consigam aproximar-se um bocado mais da política pois, em Portugal como em outros sítios, há muito ignorância sobre o que se passa. Isso também ocorre porque a informação é um bocado opaca e, portanto, é essa a nossa missão também.